



HAL
open science

Automated modeling of social touch in human-machine interaction

Wail El Bani

► **To cite this version:**

Wail El Bani. Automated modeling of social touch in human-machine interaction. Human-Computer Interaction [cs.HC]. Sorbonne Université, 2021. English. NNT : 2021SORUS216 . tel-03698457

HAL Id: tel-03698457

<https://theses.hal.science/tel-03698457>

Submitted on 18 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

présentée à

Sorbonne Université

par

Wail EL BANI

pour obtenir le diplôme de

Doctorat de Sorbonne Université

Spécialité : Robotique

Automated modeling of social touch in human-machine interaction

soutenance prévue le 07 Mai 2020

JURY

M. Mehdi AMMI	Professeur des Universités à l'Université de Paris 8	Rapporteur
M. Dirk HEYLEN	Professeur des Universités à l'Université de Twente	Rapporteur
Mme. Malika AUVRAY	Chargée de recherche à Sorbonne Université	Examineur
M. Gilles BAILLY	Chargé de recherche à Sorbonne Université	Examineur
Mme. Indira THOUVENIN	Professeure des Universités à l'UTC	Examineur
M. Mohamed CHETOUANI	Professeur des Universités à Sorbonne Université	Directeur de thèse

Résumé

Le toucher est le premier sens à se développer et le premier moyen de contact avec le monde extérieur. Le toucher joue également un rôle clé dans la communication socio-émotionnelle : nous l'utilisons pour communiquer nos sentiments, susciter des émotions fortes chez les autres et influencer leur comportement. Malgré son importance, le toucher est une modalité sous-étudiée dans l'interaction humain-machine (IHM) et l'informatique affective par rapport à l'audition et la vision. Lorsqu'ils interagissent avec des machines, les humains recherchent et s'attendent à une interaction tactile, et par conséquent la conception de systèmes capables d'interpréter et de répondre de manière appropriée au toucher améliorerait la qualité de l'interaction entre les humains et les machines. Dans cette thèse, nous avons travaillé sur l'amélioration des systèmes de reconnaissance tactile de la machine.

Avant de concevoir ces systèmes, nous avons d'abord introduit les mécanismes par lesquels le sens du toucher traite les stimuli physiques puis on a discuté d'un point de vue socio-psychologique des facteurs qui influencent le comportement tactile et son interprétation dans l'interaction sociale; puis on a passé en revue la littérature sur la reconnaissance automatique du toucher social. La construction de ces systèmes nécessite principalement de relever deux défis : (i) la détection et la collecte de données; (ii) extraction de caractéristiques du signal et la reconnaissance tactile automatisée.

La plupart des systèmes de reconnaissance tactile sociale nécessitent une étape d'extraction de caractéristiques (*features engineering*), ce qui les rend difficiles à comparer et à généraliser à d'autres bases de données. Dans cette thèse, nous proposons une approche bout à bout (*End-to-End*) basé sur l'attention pour la reconnaissance des gestes tactiles, qu'on a évalué sur deux bases de données publics (CoST et HAART) dans le contexte du ICMI 15 *Social Touch Challenge*. Notre modèle a donné un niveau de précision similaire à l'état de l'art (61% pour CoST et 68% pour HAART) et utilise l'auto-attention (*self-attention*) comme alternative au *features engineering* et aux réseaux de neurones récurrents.

Ensuite, nous introduisons une nouvelle façon de collecter des données tactiles socio-affectives, un cadre que nous avons nommé stimuli-imitation (SI). Les méthodes précédentes utilisaient des étiquettes affectives (par exemple, l'amour) ou des étiquettes de gestes tactiles (par exemple, câlin) pour susciter des gestes tactiles socio-affectifs. Ces stimuli sont vagues, séparés du contexte social et introduisent une forte variabilité inter-sujets. Le but de notre méthode est double : (i) fournir un contexte social en passant d'un stimuli basé sur une étiquette à un stimuli basé sur un scénario; et réduire la variabilité entre les sujets en contraignant la sollicitation des gestes tactiles à l'aide de l'imitation. Dans notre étude, nous avons examiné dans quelle mesure

l'imitation produit des gestes tactiles cohérents ; et aussi l'influence du contexte social introduit par les scénarios sur l'élicitation des affections. Enfin, nous avons développé un système de reconnaissance d'affect personnalisé qui utilise des données sur le toucher et les traits individuels (scores d'empathie et de personnalité) pour reconnaître l'état émotionnel des participants au-delà des niveaux de chance ; et un système qui reconnaît les traits individuels à partir des données tactiles.

Mots clefs : toucher social, reconnaissance de gestes tactile, imitation, informatique affective, collect de données, interaction human-machine.

Abstract

Touch is the earliest sense to develop and the first mean of contact with the external world. Touch also plays a key role in our socio-emotional communication : we use it to communicate our feelings, elicit strong emotions in others and modulates behavior (e.g. compliance). Although its relevance, touch is an understudied modality in Human-Machine Interaction (HMI) and Affective Computing (AC) compared to audition and vision. When interacting with machines, humans seek and expect tactile interaction thus designing systems able to interpret and respond appropriately to touch would enhance the quality of interaction between humans and machines.

In this thesis we worked toward improving the machine *touch recognition* part of the interaction loop. Before designing these systems, we first introduced the mechanisms by which the sense of touch process physical stimuli (bottom-up) then discussed from a socio-psychological standpoint the factors that influence touch behavior and it's interpretation (top-down) in social interaction; then reviewed the literature on automatic Social Touch recognition. Building these systems requires mainly facing two challenges : (i) sensing and data collection ; (ii) features extraction and automated touch recognition. Most of the social touch recognition systems require a feature engineering step making them difficult to compare and to generalize to other databases.

In this thesis, we propose an end-to-end approach. We present an attention-based end-to-end model for touch gesture recognition evaluated on two public datasets (CoST and HAART) in the context of the ICMI 15 Social Touch Challenge. Our model gave a similar level of accuracy : 61% for CoST and 68% for HAART and uses self-attention as an alternative to feature engineering and Recurrent Neural Networks. Next, we introduce a novel way for collecting socio-affective touch data, a framework that we named *stimuli-imitation (SI)*. Previous methods employed affective labels (e.g Love) or touch gesture labels (e.g Hug) to elicit socio-affective touch gestures. These stimuli are vague, divorced from social context and introduces high inter-subject variability. The purpose of our method is two fold : (i) provide a social context by transitioning from a label-based to a scenario-based stimuli ; and reduce variability between subjects by constraining touch gesture elicitation using imitation. In our study we investigated to which extent imitation produce consistent touch gestures ; and also the influence of social context introduced by scenarios on affect elicitation. Lastly, we developed a personalized affect recognition system that uses touch and individual traits data (empathy and personality scores) to recognize the emotional state of participants beyond chance levels, and a system that recognize individual traits from touch data.

Keywords : social touch, touch gesture recognition, imitation, affective computing, data collection, Human-Machine interaction.

Table of contents

List of Figures	x
List of Tables	xi
1 Introduction	1
1.1 Context	1
1.2 Motivation	2
1.3 Research Goals	2
1.4 Main contributions	3
1.5 Dissertation roadmap	4
2 Social Touch : background and state-of-the-art	5
2.1 Introduction	5
2.2 Touch	5
2.2.1 Touch and Human development	5
2.2.2 The Sense of Touch	6
The Discriminative Touch System	6
The Affective Touch System	6
2.3 Social Touch	7
2.3.1 Social Touch Modulators	7
2.3.2 Interpretation of Social Touch	8
2.3.3 Effects of Social Touch	9
Emotion Communication	9
Behavior and Attitude Modulation	9
Physiological and Emotional well-being	10
2.4 Social Touch in Human-Machine Interaction	11
2.4.1 Mediated Social Touch	11
2.4.2 Sensing and Automatic Recognition of Social Touch	12
Affective Computing & Touch	13
2.4.3 Touch Elicitation approaches	17
2.4.4 Applications	17
3 Touch Recognition with Attentive End-to-End Model	19
3.1 Introduction	19

3.2	Related work	20
3.3	Social touch challenge datasets and protocol	21
3.3.1	CoST : Corpus of social touch	21
3.3.2	HAART : Human-Animal Affective Robot Touch	21
3.4	Attentive Touch Model	22
3.4.1	Intra-Attention	22
	Multi-Head Self-Attention :	22
3.4.2	Spatio-temporal encoding	24
	Spatial Encoding :	24
	Positional Encoding :	24
3.4.3	Inter-attention	25
3.5	Touch gesture recognition	26
3.5.1	Data Preprocessing	26
3.5.2	Hyperparameters Tuning and training	27
3.6	Results and Discussion	28
3.7	Conclusion	30
4	<i>Stimuli-Imitation : Elicitation Approach for Social Touch Gesture Data Col- lection</i>	31
4.1	Introduction	31
4.2	Stimuli-imitation framework	32
4.2.1	Socio-emotional encoding	33
4.2.2	Imitation	33
4.2.3	Framework Evaluation	33
4.3	User study	36
4.3.1	The Mannequin	36
4.3.2	Stimuli selection & Sensors placement	36
4.3.3	Study setup	41
4.3.4	Participants	41
4.3.5	Individual traits assesment	42
	Interpersonal Reactivity Index (IRI)	42
	Big Five (BF)	43
4.3.6	Participants self-assessment of Valence and Arousal	43
4.3.7	Experiment protocol	44
4.4	Social Touch Recognition	45
4.4.1	Train and Test set	46
4.4.2	Feature Extraction	47
4.4.3	Model and Training	47
4.4.4	Results	48
	Touch gesture and SATg recognition	48
	Cross-Comparing Models	48

4.5	Affect Elicitation through observed Socio-Affective Touch	50
4.5.1	Annotations consistency	51
4.5.2	Influence of Context on Valence and Arousal annotations	52
4.5.3	Interactions with Individual Traits : Empathy and Personality	52
4.6	Discussion	54
4.7	Conclusion	56
5	<i>Recognition of Affect and Individual Traits through Touch</i>	59
5.1	Introduction	59
5.2	Methodology	60
5.2.1	Collected data	60
	Socio-affective Touch data	60
	Individual Traits data	61
5.2.2	Labels	61
	Emotional labels	61
	Individual Traits labels	62
5.2.3	Loss functions	63
5.3	Analysis and discussion	64
5.3.1	Training	64
5.3.2	Personalized Affect Recognition	64
	Influence of Individual Traits on the classifier performance	64
	Evaluation of RF_{X+IT} using loss functions	66
5.3.3	Individual Traits recognition from Touch	70
5.4	Conclusion	72
6	<i>Conclusion and Perspectives</i>	73
	Bibliography	77
A	Questionnaires	93
A.1	Big Five Inventory francais (BFI-Fr)	93
A.2	Index de Réactivité Interpersonnel (IRI-FR)	95
B	Valence and Arousal recognition	99
B.1	Arousal and Valence recognition	99
B.1.1	Arousal recognition	99
B.1.2	Valence recognition	100

List of Figures

1.1	Thesis challenges and contributions.	3
2.1	Interpretation of touch in a Human-Human interaction and it's effects.	11
2.2	Touch surfaces examples	13
2.3	Affect recognition pipeline : (i) Sensing ; (ii) automatic recognition (one step or two step system) ; (iii) affect model (e.g Russel circumplex) . . .	15
2.4	Paro Seal Robot	18
3.1	Natural language processing example of self-attention applied to a sequence of two words. See Jay Alammar <i>blog</i>	23
3.2	Positional encoding used for CoST (sequence-length = 600 ; d= 64). For a position i we add a row i from this matrix to the CNN output vector (d=64) to encode position.	25
3.3	Attentive Touch Model : architecture	26
3.4	Confusion matrix for HAART dataset	29
3.5	Confusion matrix for CoST dataset	29
4.1	Stimuli-imitation protocol : (i) participant read a scenario and watch the associated video (stimulus) taken from SATED [89] ; (ii) Evaluate in term of valence and arousal the stimulus ; (iii) imitate the observed touch gesture on the mannequin.	34
4.2	Scenario protocol : (i) participant read scenario taken from SATED [89] ; (ii) participant recall from past experience or use his imagination to produce the touch gesture.	35
4.3	This figure shows 13 stimuli taken from SATED [89]. Each stimuli is composed of 3 socio-affective touch expression videos varying in valence and arousal.	37
4.4	The figure shows the annotation form and the body-part map used by annotators	37
4.5	For each selected stimulus (3 videos) six annotators indicated where the touch-recipient was touched. For instance, in the Tap1_neu stimulus the shoulder was the only body-part touched.	38
4.6	Participant hugging the mannequin.	39

4.7	The matrix sensor (right side) render position and pressure data when touched. Four matrix sensors are placed on the mannequin left arm, chest, left shoulder and back. The purpose of the mannequin-sensor is to collect social touch data from various regions of the body simultaneously.	41
4.8	Experiment setup	42
4.9	Self-assessment Manikin [17]	43
4.10	Stimuli-imitation tasks.	45
4.11	Scenario task.	45
4.12	Difference between an SI and S stimulus in the case of a hug touch gesture.	46
4.13	Touch gesture and SATg recognition.	49
4.14	Difference between an SI and SATED stimulus in the case of a Hug touch gesture	50
4.15	Inter-subject consistency	51
4.16	Intra-subject consistency	52
4.17	Valence annotations for each stimuli in the SI and SATED conditions.	53
4.18	A representation of Stimuli-imitation and Scenario prototypes.	56
5.1	Personalized Affect Recognition	60
5.2	Location of emotional labels on the Affect-grid [4]	62
5.3	Histogram of emotional labels (e_i) obtained from the valence and arousal scores.	62
5.4	Random Forest features importance result.	65
5.5	Emotional labels confusion matrix.	68
5.6	Percentage decrease in loss $p(D)$ on each emotion.	68
5.7	Near misses rate for each emotion.	69
B.1	Arousal confusion matrix.	99
B.2	Valence confusion matrix.	100

List of Tables

2.1	Social Touch Recognition. The design of experiments are : Between-subject (B), Within-subject (W) and Subject-independent (SI)	16
3.1	Attentive Touch Model : Hyper-parameters (d_{model} choice is dictated by the CNN hyper-parameters)	22
3.2	Evaluation : Baselines and Attentive Touch Model	27
4.1	Selected stimuli.	40
4.2	Cross-Comparing Touch Gesture models using accuracy scores.	49
4.3	Cross-Comparing Socio-Affective Touch Gesture models using accuracy scores.	49
4.4	Influence of empathy and personality on Arousal.	54
4.5	Influence of empathy and personality on Valence.	54
5.1	Mean and standard deviation of Personality dimensions and Empathy subscales.	61
5.2	Euclidean loss for e_1	63
5.3	Classification results of emotional labels (e_i) reported using Mean and Std of f1-scores obtained from the 10fold cross validation for different conditions. At the bottom of the table results were summarized using Accuracy and f1scores average.	66
5.4	Percentage decrease in loss for the Euclidean and Discrete loss functions calculated for each emotion	67
5.5	Empathy classification results reported with f1-scores.	70
5.6	Personality classification results reported with f1-scores.	71
A.1	Mean and std of personality traits	95
A.2	Mean and std of Empathy	98
B.1	Arousal classification results for RF_X , RF_{X+IT} and baselines. The metric used are f1-scores, f1-scores average and accuracy.	100
B.2	Valence classification results for RF_X , RF_{X+IT} and baselines. The metric used are f1-scores, f1-scores average and accuracy.	101

Chapter 1

Introduction

1.1 Context

Touch is the earliest sense to develop [12] and the first mean of communication between parents and newborns [97]. It is a powerful force in the human development, shaping learning, attachment, communication and the formation of socioemotional bonds from infancy and throughout life [24]. In interpersonal communication, touch elicits and regulates human emotion [65], and alone can communicate distinct emotions [68]. In addition to communicating and eliciting emotions, touch influences people attitudes and behaviors : a touch on the arm or shoulder by a waitress -in a restaurant- increases the size of the tip regardless of customer gender [33] (Midas touch effect). Touch also contributes to our physical well being [46] : documented positive physiological effect of touch includes decrease in blood pressure, heart rate, cortisol levels (stress hormone) and increase in oxytocin levels (love hormone) [63] [61].

Historically, research on touch was conducted by psychologist, that studied the interaction of humans with objects (the sensing and manipulation by touch). The field was referred to as *haptics* derived from the word "haptesthai" which means "to touch" in Greek [128]. In the last century, technological breakthroughs made machine omnipresent in our daily life, serving various applications such as gaming (e.g controllers), communication (e.g touch screens) and health (e.g tele-surgery) [39]. The word haptics was therefore redefined to include the interaction between humans and machines and was mainly focused on the discriminative aspect of the human sense of touch. Nonetheless, the sense of touch has the ability to induce feelings ranging from unpleasant (e.g being pinched) to pleasant (e.g being caressed) and is important for the establishment and maintenance of affiliative bonds between humans [132]. The discovery of CT-afferent [94], a class of peripheral nerve receptors that are useless for discriminative touch (receptor with low spatial and temporal resolution), and that responds strongly to slow stroking (similar to how humans stroke) led researchers to propose a social touch hypothesis. A hypothesis claiming that there is a

dedicated pathway for processing affective content of touch [73]. Both of these systems (discriminative and affective) work together to create an affective percept of touch thus providing a channel of emotional communication and elicitation [66] that guides Humans through the social world. In short, the sense of touch play a pivotal role in affective communication. A novel research area called *affective haptics* took interest into designing devices and systems that can detect, process and display the emotional state of a humans [39] through touch. Incorporating these systems into Human-Machine interaction is crucial for developing efficient interactive systems.

1.2 Motivation

Touch plays an important role in human interaction, it increases trust, worthiness, warmth, social presence and emotional connection [77][103]. It also provides us with an intimate and private channel of emotional communication[66]. Although touch is crucial to our specie, it has received little attention from the affective computing (AC) and Human-Machine interaction (HMI) communities compared to audition and vision [22]. Building machines able to detect and interpret socio-affective touch would benefit number of domains including human-robot interaction [5] [6] where therapeutic robots –that relates to the human affect– such as Paro [121] could be used in nursing homes to improve the socialization of the elderly, and in hospitals to reduce patients stress level. In this thesis, we work towards integrating touch into interactive systems to leverage communicative and emotional channels between humans and machines.

1.3 Research Goals

A socially intelligent machine should be able to sense, recognize and interpret social touch signals in order to respond appropriately to human touch behavior (Interactive cycle of intelligent agent [80]). In this dissertation we will focus on the first three modules (sensing, recognition and interpretation). We believe that the integration of socio-affective touch into interactive systems requires facing two main challenges : (i) Sensing & data collection ; (ii) Automated recognition. The figure 1.1 maps the aforementioned challenges to the dissertation contributions discussed in more details in the next section.

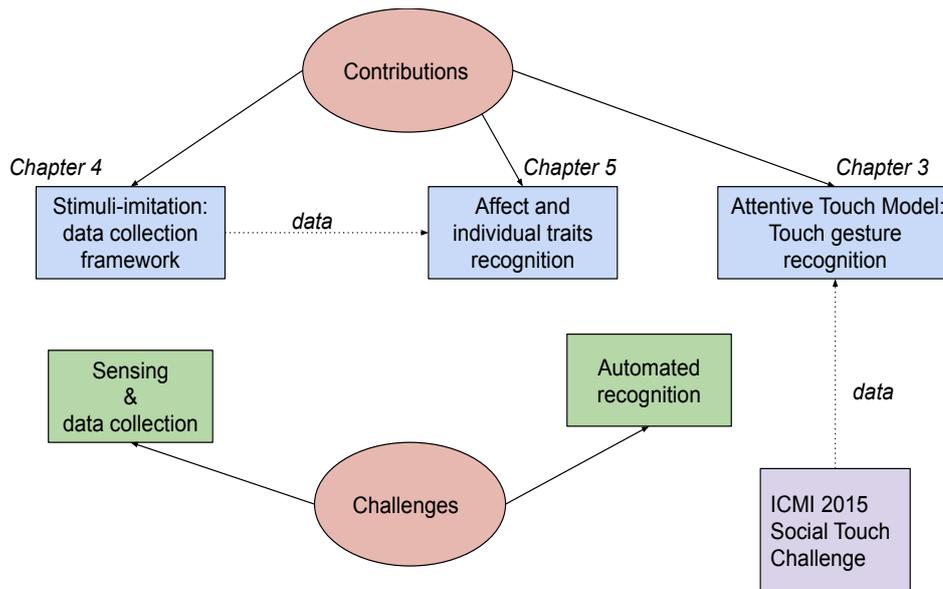


FIGURE 1.1 – Thesis challenges and contributions.

1.4 Main contributions

The main contributions of this dissertation are :

We propose an end-to-end approach for touch gesture recognition (chapter 3) : Most of the social touch recognition systems require a feature engineering step making them difficult to compare and to generalize to other databases. In this chapter, We present an attention based end-to-end model for touch gesture recognition evaluated on two public datasets (CoST and HAART) in the context of the ICMI 15 Social Touch Challenge. Our model gave a similar level of accuracy : 61% for CoST and 68% for HAART and uses self-attention as an alternative to feature engineering and Recurrent Neural Networks.

We introduce a novel way for collecting socio-affective touch data, a framework that we named stimuli-imitation (chapter 4) : Previous methods employed affective labels (e.g Anger) or touch gesture labels (e.g Hug) to elicit socio-affective touch gestures. These stimuli are vague, divorced from social context and introduces high inter-subject variability. The purpose of our method is two fold : (i) provide a social context by transitioning from a label-based to a scenario-based stimuli; and reduce variability between subjects by constraining touch gesture elicitation using imitation. In our study we investigated to which extent imitation produce consistent touch gestures; and also the influence of social context introduced by scenarios on affect elicitation.

Recognition of affect and individual traits (chapter 5) : We show that we could recognize affective content in social touch using data collected through stimuli-imitation

framework, and that adding individual traits such as empathy and personality as input enhanced the affect classification. Furthermore, we also showed that it is possible to recognize individual traits (e.g High perspective taking) from touch data only.

1.5 Dissertation roadmap

The dissertation will begin with chapter 2, an introductory chapter in which we present the role of touch from a developmental, physical and socio psychological perspective. The rational is to understand how touch is interpreted in a human to human interaction in order to design useful touch interpretation systems for Human-machine interaction. In chapter 3, we will develop an End-to-End touch gesture recognition deep learning model on the CoST and HAART data sets following the ICMI 15 Social touch challenge protocol. In chapter 4, we will introduce a novel framework for collecting socio-affective touch gestures that we named stimuli-imitation in which subjects imitate videos of socio-affective touch interactions performed by actors. In chapter 5, we will develop an affect recognition and individual traits recognition models using data collected from the stimuli-imitation framework. In chapter 6, we will conclude the dissertation by putting together all our findings and by presenting future works.

Chapter 2

Social Touch : background and state-of-the-art

2.1 Introduction

The purpose of this thesis is to develop Socially intelligent systems capable of interacting with humans via touch. Before delving into designing such systems, it is crucial to understand what touch represent for humans, and how it is employed in social interaction. In this chapter we review the literature on touch in Human-Human interaction and Human-Machine interaction.

2.2 Touch

In this section we discuss the importance of touch in human development and present the systems that constitute it.

2.2.1 Touch and Human development

Touch is the earliest sense to develop [12] and the first mean of communication between parents and newborns [97]. In the womb, a fetus is constantly being stimulated by the amniotic fluid and the uterus walls, preparing him for postnatal functioning (e.g breathing). The newborn continues to receive tactile stimulation by being cuddled and breastfed [98]. During infancy, parental touch is essential for the development of the social brain [24] and a shortage of it leads to cognitive, social and neurodevelopmental delays [92] [27]. Furthermore, mother-infant touch interaction is associated with a reduction of behavioral [129] and physiological [45] stress response for the infant. As they grow up, infants gain knowledge of the world around them by seeing, reaching, grasping and manipulating objects of different sizes and shapes [117][31] thus creating a mental representation by integrating visual cues and tactile sensations. As the infants become toddlers, a transition characterized by an increase

in mobility, parent-child interaction becomes dynamic thus including complex touch such as tickling and grooming [37]. Grooming for instance plays an important role in maintaining social relationships, specially the one involving pleasant sensations (e.g light stroke) [94]. Also, a naturalistic study showed that touch helped to prolong children's expressions of positive emotion [9] thus promoting positive socialization of emotions. At school, children's touch circle expands to include teachers [90] and peers [144]. And that touch patterns were determined by other factors such as age, sex and race from kindergarten to high school [145].

2.2.2 The Sense of Touch

The human skin is the largest and oldest sense organ [97], an average adult skin covers approximately $1.7m^2$ and weights around 3.6 kilograms. It contains several different sensory receptors that activates depending on the nature of stimuli. There are two systems that allow us to sense and perceive physical stimuli : (i) the *discriminative touch system* ; and (ii) the *affective touch system* :

The Discriminative Touch System

The Discriminative touch system consist of two senses : (i) the *kinesthetic sense* ; and (ii) the *cutaneous sense*. The kinesthetic sense provide information about the position of the body and limbs in space and time. It is a self-oriented sense (as opposed to the other senses) that focuses on internal events by means of receptors ingrained in the tendons, muscles and joints (*proprioceptors*) thus guiding our movement in daily activities (e.g walking, running). The cutaneous sense provide information about physical stimulation of the skin owing to various receptors. The *mechanoreceptors*, for instance, detects skin deformation (e.g pressure, vibration) and enables us to handle object and feel their shape, size and texture ; *nociceptors* responds to pain (potentially damaging stimuli) ; and *thermoreceptors* detect temperature difference between the skin and an external element (e.g air). Signals from these receptors are combined then processed by the somatosensory cortices of the brain allowing perception and interpretation of physical stimuli such as social touch.

The Affective Touch System

Whether it is an anxiety inducing nudge from a stranger or a comforting hug from a spouse, touch elicit strong emotions. The skin contains receptors that elicit emotional response ranging from negative affect (painful touch) [7] to positive affect (arousing touch) [147][86]. We experience more pleasantness by being touched than touching [137]. Also, it was shown that pleasantness is at it highest when touch occur at skin temperature [2] and at CT-optimal speeds [91] which is similar to how humans naturally stroke each other [32]. Researchers hypothesized that a class of peripheral

nerve receptors -found in hairy skin- called CT-afferent [94] are responsible for the social and affective perception of touch. However, the CT-afferent responds only to gentle stroking (CT-optimal) in hairy regions which do not explain the effect of CT non-optimal touches (e.g hand shaking). Both affective and discriminative systems play a role [95]. Furthermore, the perceived pleasantness of a touch depends also on the identity of the toucher [54] and other factors -described in the next section- where cognition is central.

2.3 Social Touch

In the last section we saw that the discriminative system works in conjunction with the affective system to produce a pleasant or unpleasant precept. In this section we enumerate extraneous factors (e.g culture, context) that modulates interpersonal touch behavior, and discuss their role in the interpretation of touch (e.g support, love). Furthermore, we also present various effect of interpersonal touch (e.g physiological well-being).

2.3.1 Social Touch Modulators

In social interactions, touch behavior patterns are influenced by several elements such as : sex, relationship, context, culture, personality and type of touch.

In his study of body accessibility - readiness of a person to permit others to contact his body [78], Jourard administered questionnaires (recall-based method) to unmarried college students (168 male, 140 female) to determine which body parts are touched or being touched by the mother, father, same-sex friends and opposite-sex friends. The results revealed a gender influence on touch patterns : females are perceived more accessible to touch than males, and that mothers and opposite-sex friends (intimate friends) were reported as doing most of the touching. Additionally, Heslin and Nguyen [70] found -using also questionnaires- that the meaning attributed to a touch is related to acquaintanceship for women and to sex for men. Put differently, men find unpleasant same-sex touch and pleasant opposite-sex touch while women find unpleasant a touch from a stranger and pleasant a touch from a close-friend. This conclusion could be easily verified in crowded places such as the Bus or Metro.

Certain settings plays an important role in inhibiting or encouraging touch behaviors. At the airport, it has been shown that approximately 60% of the observed interaction included interpersonal touch; another study reported a touch percentage of 83% for greeting [55]. Similarly, it was observed that pre-school children's engaged more in interpersonal touch when they were playing outside than inside [143]. On the other hand, Remland et al. [110] found that, in several European locations,

touch occurred only in 9% of the interactions in public places. However, when it comes to sport events, touch behavior tend to be frequent between players compared to a normal social interaction [127].

Cultural difference also influences touch behavior, some cultures are viewed as more tactile than others. Remland et al. [109] found that Greeks and Italians touched more frequently than Dutch, French and English in public places. Similary, Jourard [78] observed the frequency of touch in four different countries. He found that touch occurred at cafe tables more frequently in Puerto Rico and Paris than England or the USA. Latin Americans manifest a high level of touch in public places. When interacting with friends and family they employ at least a hug or a handshake [93][122]. In contrast to Latin America, Asian cultures are avoidant of physical contact, for instance, Japanese uses a bow for greeting rather than a handshake [13].

Attitude towards touch varies also within the same culture. People with different personality respond differently to touch. In [142], the authors found that socially anxious people avoid situation that involves touch and were less incline to pursue or initiate touch behavior in public. Also, touch-avoiders compared to touch-seekers exhibit low self-esteem and signals low social status [20]. Furthermore, Dorros et al. [36] studied the relationship between touch perception and the big five personality dimensions. The result showed that agreeableness and openness were significant predictor of positive perception of touch than gender or relationship satisfaction with the toucher (measured by Relationship Assessment Scale [62]).

2.3.2 Interpretation of Social Touch

Before the classical study on the meaning of touch conducted by Jones and Yarbrough (1985) [77], most research on social touch used a recall-based approach (e.g self report) and was concerned with who touches whom, where, and how often. In few cases where the meaning of touch was examined, such as the study of Hesling et al. (1983) [70] where respondents filled a questionnaire answering the question : *"What does it mean (e.g playfulness) to me when a close person (or stranger) of the opposite sex (or same sex) touches the indicated area of my body (e.g arm) in a certain manner (e.g squeeze)?"* the role of context in shaping the meaning of a touch was ignored. In their study, Jones et al. [77] considered a naturalistic approach for examining how touch behavior and contextual factors are combined to produce meaning. 39 males and females were recruited and trained to record details about their touch experience over a period of 3 days thus resulting in 1500 collected touch gestures. Qualitative analysis on data revealed 12 distinct and unambiguous meanings : support, appreciation, inclusion, sexual interest or intent, affection, playful affection, playful aggression, compliance, attention-getting, announcing a response, departure, and greeting. There was also touch with Hybrid meaning such as greeting/affection and

departure/affection, and other ambiguous touches, for instance, accidental touch and touch that refer to appearance. Furthermore, this study refutes the "universality of touch gesture meaning" in favor of a context-dependent meaning of touch in which context is defined as a set of verbal, nonverbal and situational factors that accompany a given touch gesture and clarify its meaning. As an example, a touch in the arm could be interpreted as communicating support (e.g doctor touching patient in a hospital) or attention getting (e.g asking a stranger for direction) depending on the context, different touch gestures (e.g pat and holding) could convey the same meaning (e.g support) in a specific context.

2.3.3 Effects of Social Touch

Emotion Communication

In Burgoon et al. [19] it was found that touch convey more trust, affection and receptivity than its absence. Touch can communicate positive social messages such as support, appreciation and sexual attraction or negative ones like anger and frustration. Beside conveying complex social messages, touch can communicate distinct discrete emotions. Hertenstein et al. [67] conducted a study where two strangers separated by an opaque barrier interacted with each other using only touch on the arm. An encoder was asked to convey 12 emotions to a decoder and the decoder had to guess which emotion the encoder was communicating. The results show that, except for Happiness and surprise, Prosocial Emotions (Love, Gratitude, Sympathy) and Ekman's Emotions (Anger, Fear, Happiness, Sadness, Disgust, Surprise) were recognized at above-chance level; while Self-focused emotions (Embarrassment, Envy, Pride) were not recognized at above-chance level. In [66], Hertenstein et al. conducted an unconstrained similar study where participants could touch anywhere on the body that was appropriate. In this study 8 emotions were involved. The result showed that Happiness and Surprise were reliably communicated as opposed to the first study. While strangers in the aforementioned studies were not able to communicate self-focused Emotions through touch. Thompson and Hampton [136] showed that romantic couples were able to communicate envy and pride at above-chance level. Therefore the relationship between the two people and the body-parts involved play a role in encoding and decoding of emotions when using touch.

Behavior and Attitude Modulation

Touch influences people's attitudes toward persons and places. Students that received a casual touch (touch condition) by a library clerk while returning a library card evaluated positively the clerk and the library compared to the No-touch condition

[49]. While in this study the clerk was evaluated on professional attributes (e.g helpfulness), similar results were found on the evaluation of a sales-man personal attributes (e.g friendliness) [42]. Moreover, it was demonstrated in several naturalistic studies that touch modulates people's compliance towards a variety of requests. For example, a touch on the arm or shoulder by a waitress in a restaurant increases the size of the tip regardless of customer gender [33], and students that were touched by their teacher volunteered more in class [57]. Touch have a powerful effect on people. In Gallace and Spencer [52], three explanations as to why touch has a powerful effect on people were provided : (i) The perception of great need and positive feeling increases compliance rate. The recipient of touch assumes that the toucher is in genuine need and that the toucher likes and trusts them; (ii) Positive response to touch in adulthood originate from the early association between touch and stress reduction in early life; (iii) Touch itself generates pleasant sensations through receptors in the skin [95] thus eliciting compliance.

Physiological and Emotional well-being

Touch contributes to our physical well being [46] : documented positive physiological effect of touch includes decrease in blood pressure, heart rate, cortisol levels (stress hormone) and increase in oxytocin levels (love hormone) [63] [61]. *Massage therapy* is an effective form of touch used for treating pain, depression, stress and autoimmune conditions (e.g asthma) [46]. Massage therapy also enhances attentiveness and alertness [47], Specially for touch aversive children with ADHD [48] (attention deficit hyperactivity disorder); and improves sleep for children Autism [44].

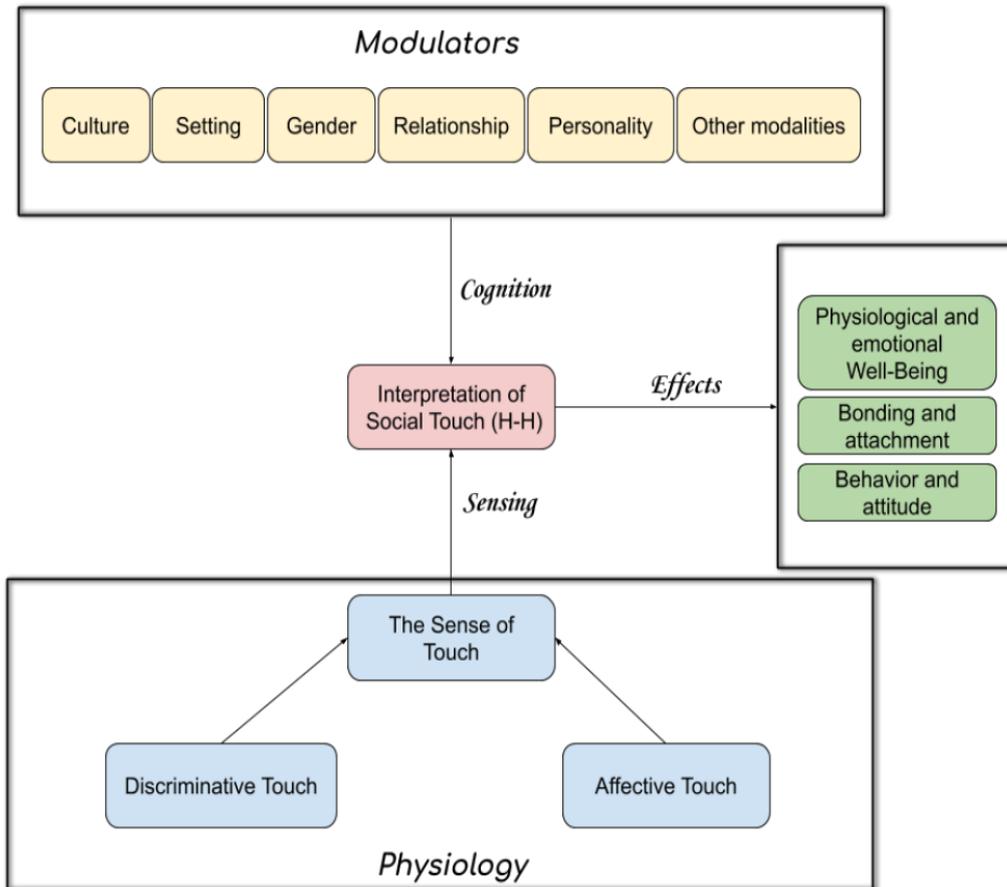


FIGURE 2.1 – Interpretation of touch in a Human-Human interaction and its effects.

2.4 Social Touch in Human-Machine Interaction

The study of Touch in Human-Machine interaction is at the crossroad of *Social signal processing*, a field that aims at modeling, analysing and synthesizing of social signals (observable behaviors that people display in social interaction), and *affective computing* a field that specializes in developing systems capable of recognizing, interpreting and simulating human affect [106]. In the previous sections we saw how touch fits in a Human-Human interaction, and presented its various effects. In this section we review the literature on Touch in Human-Machine interaction.

2.4.1 Mediated Social Touch

In the 21st century, interpersonal communication occurs more often in a mediated fashion. Existing devices rely mainly on audition and vision, and rarely support touch. Recent advances in haptic technology, inspired researchers to explore how haptic devices could be used for mediated social communication.

Mediated social touch technology role is to enable haptic communication between people who are physically apart using tactile or kinesthetic feedback [59]. Unlike actual touch, mediated touch is less sensorially rich, can happen asynchronously, does not require physical proximity. In spite of these differences, research on mediated touch reported similar results to actual touch in terms of emotion elicitation, emotion communication and behavior modulation. In [138] findings show that mediated touch through air jet elicit distinct feelings. In [115] was found that emotions could be conveyed by a phone through Vibrotactile Rendering. Furthermore, a study found that participants that received a vibrotactile feedback during a chat session complied more when asked to pick up a coin (virtual midas touch) [60]. Moreover, mediate affective touch reduces heart rate (physiological effect) [21]. And increases intimacy, sympathy, connectedness towards the remote toucher (psychological effect) [135, 140]. For an exhaustive review we refer the reader to Huisman survey [73] and Haans review [59].

2.4.2 Sensing and Automatic Recognition of Social Touch

When interacting with humans, machines are expected to recognize and respond appropriately to touch [88]. Especially in ways that relate to the state of the human [6]. Previous research included touch recognition of social messages (e.g Greeting), emotions (e.g happiness) and touch gestures (e.g grab) using a combination of features extraction techniques and machine learning algorithms (e.g Random Forest). Different approaches were considered to sense and interpret human touch (table 2.1). These approaches vary in terms of touch sensor, touch surface, experimental design (subject-independent and between/within-subject) and were evaluated on classification accuracy.

Different types of sensors were used to detect : contact [84] (e.g charge-transfert sensor), force/pressure [82] (e.g piezoresistive sensor), temperature difference [130] (e.g thermistors), inertia [29] (e.g gyro sensor), proximity [130] (electric-field sensor) and light dynamic touch such as a *stroke* [50] (e.g conductive fur sensor). These sensors were placed on a human-friendly touch surfaces to simulate, for instance, a human [29] (Humanoid), a human body-part [125] (mannequin arm), a pet [4] (haptic creature) or a toy [87] (Huggable) thus creating suitable interfaces for social touch interactions. In social touch studies data was gathered from multiple subjects to train subject-independent models, and from a single subject for prototyping and testing. Furthermore, several feature extraction techniques were developed to enhance the classification of the models. The ICMI social touch challenge [81], described in detail in the next chapter, demonstrated the Human Machine Interaction (HMI) community efforts towards developing robust features and accurate models for social touch gestures recognition. Touch gesture recognition results showed a significantly higher

2.4. Social Touch in Human-Machine Interaction

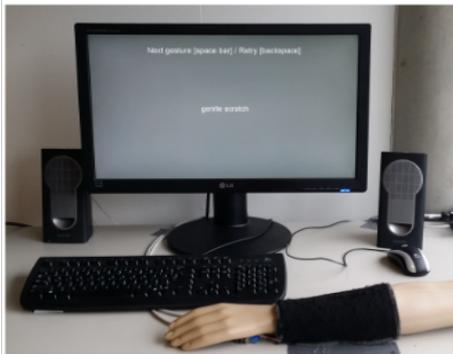
accuracy scores for within-subject than between-subject designs, even though subjects were exposed to the same stimuli which is usually a word (e.g rub) or a touch dictionary (rub : *Move your hand repeatedly back and forth on the arm with firm pressure*).



Haptic Creature (2008)



Huggable (2006)



Mannequin arm (2014)



Kirin (2012)

FIGURE 2.2 – Touch surfaces examples

Affective Computing & Touch

Research on automatic interpretation of affective touch originating from a human, included two approaches (Figure 2.3) : (i) one-step system; and (ii) two-step system. The one-step system derive affective meaning directly from touch sensors [124] while the two-step system identifies first the applied touch gesture then attach an affective meaning to it [4]. Both approaches aim at correctly recognizing *affect* whether it is a communicated (expressed) or an experienced affective state (felt). Affect is generally represented in a *categorical* form (e.g anger) as independent classes [40] or in a two *dimensional* space (Russel circumplex) following two axis : Valence and Arousal [113]. Although touch is an under-explored modality in affective computing, it follows similar pipeline and faces similar challenges as audition and vision [106]. Developing affective computing (AC) systems requires making meticulous choices at each step of the pipeline.

At the sensing step, a valid signal should reliably discriminate different affective states when combined with an appropriate pattern recognition method. For instance, Silvera et al. [123] set the requirement for the *artificial skin* (pressure-sensitive device), to a minimum sampling rate of (20 Hz), and a spatial resolution between 10-40mm to discriminate between multiple fingers touch (e.g. tap) and an open hand touch (e.g. pat). These requirements led to an above chance-level emotion classification. Other studies deployed multiple sensors to capture information from various modalities, as an example, the Huggable [130] captured in parallel temperature, pressure and proximity.

Beside sensing, it is important to consider which affective phenomena (e.g emotion, moods, feeling) to model. Most AC research focus on the **emotion** construct. In their review [22], Calvo et al. presented six perspectives from which emotion could be studied in AC systems : Emotions as Expression, Emotions as Embodiment, Neuroscience, Emotions as Cognition and Emotions as Social Construct. While studying affective touch from the Embodiment, Neuroscience and Expression perspectives requires respectively, physiological sensors (e.g heart rate [100]), brain imaging techniques (e.g EEG [126]) and touch-centered sensors (e.g pressure sensor [124]). The cognitive and social perspectives requires a system capable of situational assessment [120] "*appraisal*" and social analysis (e.g status, power, culture).

Most of affect recognition systems developed for Touch are bottom-up, they map physical characteristics of touch (e.g pressure, temperature) to socio-affective labels (e.g touch gesture or emotion) [43]. In one hand, these systems are context-free and socially divorced [22], meaning they are trained on touch expression data collected in an environment devoid of context (situational and social). In the other hand, we know that touch interpretation is modulated by contextual factors (Figure 2.1), therefore we believe that coupling top-down information (modulators) with bottom-up models (Figure 2.3) is the next natural step for developing systems that generalize well in new contexts. In the next section we discuss *touch elicitation approaches* employed in the literature, and in Chapter 4 their limitations towards achieving this goal.

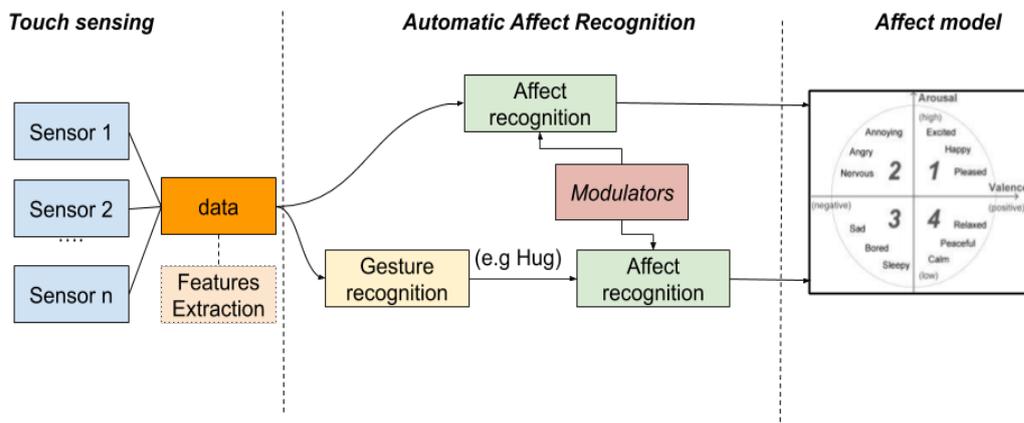


FIGURE 2.3 – Affect recognition pipeline : (i) Sensing; (ii) automatic recognition (one step or two step system); (iii) affect model (e.g Russel circumplex)

TABLE 2.1 – Social Touch Recognition. The design of experiments are : Between-subject (B), Within-subject (W) and Subject-independent (SI)

Reference	Touch surface	Sensor(s)	Touch recognition of	Acc	Design	Year
Naya et al. [101]	sensor sheet	pressure-sensitive conductive ink	5 gestures	87%	B	1999
Iwata et al. [74]	robot	force sensing resistors (FSR)	10 gestures	22-53%	W	2005
Stiehl et al. [130]	The Huggable	electric field sensor, force sensors, thermistors	8 gestures	79%	W	2005
Bailenson et al. [10]	force-feedback joystick	2d accelerometer	7 emotions	36%	B	2007
Knight et al. [87]	Huggable	pressure-sensitive conductive ink	5 gestures	20-100%	R	2009
Cooney et al. [29]	small Humanoid robot	accelerometer, gyro sensor	13 full-body gestures	77%	B	2010
Chang et al. [25]	Haptic creature	force sensing resistors	4 gestures	77%	R	2010
Kim et al. [84]	KaMERO	charge-transfer touch sensors, accelerometer	4 gestures	83%	R	2010
Ji et al. [75]	Humanoid hand	capacitive pressure sensors	4 gestures	96%	W	2011
Cooney et al. [28]	Humanoid robot "mock-up"	Microsoft Kinect and/or photo-interrupters	20 full-body gestures	82%	B	2012
Silvera-Tawil et al. [125]	mannequin arm	EIT-based sensor, force sensor	9 gestures	88%	W	2012
Flag et al. [50]	Fur	conductive fur sensor	3 gestures	82%	B	2012
Flag et al. [51]	Furry zoomorphic machine	conductive fur & piezoresistive pressure sensors	9 gestures	86/94%	B/W	2013
Nakajima et al. [99]	Emoballoon	barometric pressure sensor, microphone	6 gestures	75/84%	B/W	2013
Jung et al. [79, 82]	mannequin arm	piezoresistive fabric pressure sensors	14 gestures	54%	SI	2014
Silvera-Tawil et al. [124]	mannequin arm	EIT-based sensor, force sensor	6 emotions	32/88/47%	B/W/SI	2014
Silvera-Tawil et al. [124]	mannequin arm	EIT-based sensor, force sensor	6 social messages	51/84/50%	B/W/SI	2014
van Wingerden et al. [146]	mannequin arm	piezoresistive fabric pressure sensors	14 gestures	64%	B	2014
Altun et al. [4]	Haptic creature	force sensing resistors, accelerometer	26 gestures & 9 emotions	36/48%	B/W	2015
Sun et al. [131]	Pepper's arm	FSR matrix sensor	7 gestures	84-94%	B	2017
Zhou et al. [151]	Flat surface & dummy hand	textile pressure mapping (TPM)	7 gestures	88-93%	B	2017
Heraz et al. [64]	Smartphone	force sensitive screen	9 emotions	86-91%	B	2018
Park et al. [102]	Flat surface	charge-transfer touch sensors	4 gestures	90%	B	2019

2.4.3 Touch Elicitation approaches

Studies on automatic touch recognition were mainly interested in the classification of *touch gestures* and *emotions*. In the data collection phase of each study, subjects were given instructions on how to produce touch expressions.

To collect touch gesture data, the stimuli used was either a single word that evokes a touch expression such as a "pat", "push", "scratch" and "stroke" [74][125] or a combination of the word and its **description** such as a **pat** : *Gently and quickly touch the arm with the flat of your hand* [82]. The latter are usually taken from a carefully compiled touch dictionary. The Yohanan and MacLean touch dictionary [149], inspired by the Human-Animal interaction [14] and the Human-Human touch [141, 66] literature, contains a set of 30 touch gestures suitable for Human-Pet interaction. Several studies employed a sub-sample of the Yohanan and MacLean touch dictionary, specially the ones that used a Zoomorphic Sensor [51, 50, 23, 25], and one study adapted the dictionary to suit an interaction with a mannequin arm [82].

To collect affective touch data, subjects were given *discrete labels* that represent an emotional state such as "disgust", "anger" and "joy", a well known example is Ekman's basic emotions [40]. For instance, Bailenson et al. [10] investigated virtual emotion communication of these emotion through a 2 degree of freedom joystick, in their study participants were asked to "do your best to communicate the exact mental state to someone else who may use the joystick to attempt to identify your specific mental states from the movement at a later date." Other studies, employed *dimensional labels* taken from the *affect grid* model [114], which is a practical version of the Russel Circumplex model [113]. A prominent example, is the Altun and MacLean study [4] where participants were instructed to "**imagine** different emotional states, and to interact as if with a real pet companion" when interacting with the *Haptic creature*, rather than to communicate specific emotions. Furthermore, Silvera et al. [124] included , beside basic emotions, Social Messages borrowed from social science [69, 76, 56] such as : "attention-getting," "greeting," "acceptance," "rejection," "affection" and "animosity", then asked subjects to convey them to an artificial arm sensor. In Chapter 4 we discuss the limitation of these elicitation methods and propose a novel framework for Social Touch data that we named *Stimuli-imitation*.

2.4.4 Applications

Building machines able to detect and interpret socio-affective touch would benefit number of domains including Human-Robot Interaction [6]. Therapeutic robots that relates to the human affect such as the robotic seal Paro [121], are used in nursing homes to improve the socialization of the elderly (Figure2.4), and in hospitals

to reduce patients stress level. They receive and process tactile data in order to select which adequate behavior to display/execute. The Yohanan et al. Haptic creature [150] communicates its internal state via vibrotactile purring, ear stiffening and breath modulation; and the Stiehl et al. Huggable [130] can orient itself towards the touch, hug and nuzzle to provide affective display.

The ability to recognize affective content of touch and respond appropriately to it, enhances bonding and engagement between humans and machines; and facilitates integration in social environments, specially in the case of Human-Humamoid interaction where expectation of social intelligence is high [30]. By closing the interactive loop, Humans and Machines could establish close relationships preventing loneliness and its consequences (e.g impaired sleeping, risk of suicide, and high blood pressure).



FIGURE 2.4 – Paro Seal Robot

Chapter 3

Touch Recognition with Attentive End-to-End Model

3.1 Introduction

For machines to perceive and interpret the human touch there is a need for a reliable automatic touch recognition system. The Social Touch Challenge [81] provided two datasets CoST [82] (Corpus of Social Touch) and HAART (Human Animal Affective Robot Touch) [23] to foster research on touch gesture recognition. The systems developed in this challenge relied mainly on hand-defined feature extraction and feature selection techniques inspired from other domains such as speech recognition and image processing [133][53] [11] making them difficult to compare and to generalize to other databases.

Deep learning approaches require no feature engineering step and give similar results compared to feature-engineering based systems [71]. The state of art results were obtained using 2D-CNN for CoST [3] (author did not follow the challenge protocol) and a 3D-CNN for HAART [152]. However, there is no deep learning model in the literature that gives state of the art results for both datasets simultaneously.

This chapter presents our effort towards developing a deep learning system that generalizes well for both datasets, hence we propose an attention-based end-to-end approach to touch gesture recognition that uses self-attention as an alternative to feature engineering and Recurrent Neural Networks. Our model gave a similar level of accuracy : 61% for CoST and 68% for HAART [41].

This chapter is organized as follows : the next section will enumerate several approaches for touch gesture recognition in the context of the social touch challenge, in section 3.3 we will describe the social touch datasets and the challenge protocol, in section 3.4 we will present the *Attentive Touch Model* (ATM) architecture, in section 3.5 the training process will be presented, In section 3.6 touch gesture recognition results are presented and discussed.

3.2 Related work

In the social touch challenge, the purpose was to develop robust features and apply classification methods to improve touch gesture recognition.

In [11] authors first transformed the multidimensional signal from the pressure matrix into sixteen 1-dimensional signals that summarize the pressure (7 signals) and the region of contact (9 signals). From each signal, 19 features were extracted : mean, variance, maximum value, minimum value, median, energy, Hurst exponent, Hjorth complexity (3 features) and autoregressive model coefficients (9 features). On the total 304 features (16 signals x19 features) then a feature selection algorithm was applied (sequential floating forward search). Selected features for each datasets were used as input for a Random forest classifier.

In [53] five groups of features were considered. Sensor area distributions were calculated at frame level and gesture level for each condition : no touch, touched and touched with high pressure (Statistical distribution (SD) of pressure surface group). To capture the shape of touch gestures, each touch gesture instance was mapped to an 8 by 8 binary matrix. A cell (channel) of this matrix is set to 1 if the corresponding sensor cell reaches the high pressure value defined for each dataset (Binary Motion History group). The third set of features (Motion statistical distribution group) computed for each cell the minimum, maximum, mean, first quartile, median, third quartile, area, total variation, interquartile range, variance, skewness and kurtosis on the whole touch gesture. Authors also used a modified version of the Motion History Histogram (MHH) called Spatial Multi-scale Motion History Histogram (SMMHH) on each cell to capture the touch gestures dynamics. Finally, similar to SMMHH, Local Binary Pattern on Three Orthogonal Planes (LBPTOP) was computed [58]. Authors applied a Random forest and a Boosting algorithms on the aforementioned features.

In [134] authors divided features into three groups. A *global features* group that represent overall statistics of the gesture, for instance, the average value of pressure over all channels and all frames. A *channel-based features* group which focus on the relationship between channels (e.g average pressure of each channel over all frames). A *sequence features* group that describe the change of the gesture (e.g Fast Fourier transform based features). The concatenation of these feature groups gave a 546 feature vector. Authors then used the Random Forest model to select relevant features resulting in a 164 feature vector (see more detail in [134]). The classification algorithms applied on the feature vectors were Random Forest and Support Vector Machine.

In [72] Geometric moment (GM) and an autoencoder (AE) were used to reduce frames dimension. For each gesture, a Gaussian Hidden Markov model (GHMM)

3.3. Social touch challenge datasets and protocol

was trained on the reduced signals (GM and AE) to calculate the likelihood scores (14 for CoST and 7 for HAART) of each touch gesture. These scores represents the frame-level features. At a gesture-level, two signals were computed from the original signal : (1) signal of average pressure; (ii) signal of standard deviation. Three statistical features were calculated : average of signal (1) and (2); standard deviation of signal (2). In addition spectral information (dominant FFT frequencies) were extracted from signal (1). These feature combined were used by a logistic regression model for classification.

In [71] three deep neural networks architecture were proposed for social touch gesture recognition : (1) CNNs; (2) CNN-RNNs and (3) Autoencoder-RNNs. The CNNs and CNN-RNNs architectures used a 3D-CNN to extract spatio-temporal features on a sliding fixed windows. The CNN classified individual windows separately and assumed that a window had the same label as the global touch gesture while the CNN-RNNs fed sequentially the spatio-temporal representation from the 3D-CNN to an RNN to achieve global touch gesture recognition. The Autoencoder-RNN architecture, employed similar autoencoder to one described in [71] to reduce frames dimensionality. The reduced sequences were used as input to the RNN.

3.3 Social touch challenge datasets and protocol

3.3.1 CoST : Corpus of social touch

The CoST dataset [82] consists of 14 touch gestures (grab, hit, massage, pat, pinch, poke, press, rub, scratch, slap, squeeze, stroke, tap, and tickle) applied by 31 subjects to a pressure sensitive 8x8 grid sensor wrapped around a mannequin arm in 3 variations (gentle, normal, rough) and 6 repetitions. The grid sensor collected variable length time series sampled at 135 Hz for the 64 channels, and each channel value ranged from 0 to 1023. The dataset contains a total of 7812 touch gestures (14 touch gestures x 31 subjects x 6 repetitions x 3 variations), whereas the Social Touch Challenge [81] considered only the gentle and normal variations thus leading to 5203 touch gestures. For this challenge, 21 subjects were assigned randomly to the train set, and 10 subjects to the test set.

3.3.2 HAART : Human-Animal Affective Robot Touch

The HAART dataset [23] contains 7 touch gestures (No touch, Constant, Pat, Rub , Scratch, Stroke, Tickle) collected from 10 participants using a zoomorphic machine equipped with a 10x10 grid pressure sensor. Each touch gesture was collected for a fixed window of 10 seconds at 54 Hz, and each pixel (channel) represent a pressure value ranging from 0 to 972. Spatially, data was trimmed to 8x8 (64 channels) to

TABLE 3.1 – Attentive Touch Model : Hyper-parameters (d_{model} choice is dictated by the CNN hyper-parameters)

Module	Layer	HParams	CoST	HAART
CNN	Conv2D	Filters#	4	3
		Kernel size	3	3
Intra	MaxPool2D	Pool size	2	2
	MHSA	heads	8	8
	MHSA	d_{model}^*	64	48
	FC1	Num_units	32	3
Inter	Bahdanau_att	d_{att}	32	4
	FC2	Num_units	32	12

reduce noise coming from the edge of the sensor. Temporarily, the first and the last second were trimmed to provide an 8 seconds of continuous capture. The touch gestures were performed on 12 different conditions : 3 substrate conditions (firm and flat; foam and flat; foam and curve) and 4 fabric cover conditions (none; short minkee; long minkee; synthetic fur). HAART contains 829 gestures and each gesture is 432 frames long ($8 \times 54 = 432$). In the social touch challenge data was split into 7 participants for training and 3 participants for testing.

3.4 Attentive Touch Model

In this chapter, we propose the Attentive Touch Model (Figure 3.3), which combines attention mechanisms and learning spatio-temporal representations for touch recognition. The rationale is that learning spatial encoding is of particular interest in touch gesture recognition since signals are per definition pressures obtained at different places of the sensors. The nature and variation of touch gestures in respect to the different gestures, user styles or hardware devices negatively impact the performance of traditional sequential representation learning such as RNN [35]. The spatio-temporal representation learning is combined with intra- and inter- attention mechanisms able to optimize the learning process.

3.4.1 Intra-Attention

Multi-Head Self-Attention :

Self-attention is an attention mechanism that computes a representation of a sequence by relating different positions (time-steps) within the same sequence. The encoding of each position is computed using the Scaled Dot-Product Attention [139] defined in (3.1) where Q, V and K are respectively the Query, Value and Key (d_k -dimensional) Matrices, and X is the input. Put differently, a position is encoded using

3.4. Attentive Touch Model

a weighted sum of all positions including the encoded position itself. For example, the contribution of position i (q_i, k_i, v_i) to the encoding of the position j (q_j, k_j, v_j) is measured by $\text{Attention}(q_j, k_i, v_i)$. An example of how self-attention is calculated is provided in Figure 3.1.

Furthermore, the multi-head self-attention is a result of running the Scaled Dot-Product attention multiple times in parallel with different trainable parameters W^q , W^k and W^v (3.3). These representations (heads) are concatenated then linearly transformed (3.2) into a single representation. This idea is similar to *Ensemble Learning* where we combine different representations to improve performance.

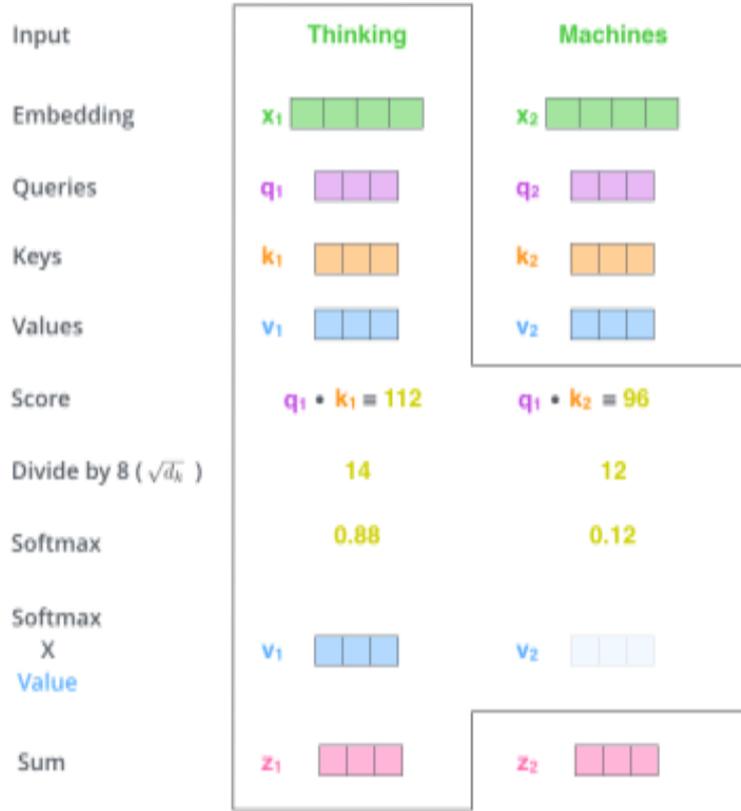


FIGURE 3.1 – Natural language processing example of self-attention applied to a sequence of two words. See Jay Alammr [blog](#)

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3.1)$$

$$Q = W^q X; K = W^k X; V = W^v X;$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1; \dots; \text{head}_h) W^0 \quad (3.2)$$

$$\text{head}_i = \text{Attention}(QW_i^q, KW_i^k, VW_i^v) \quad (3.3)$$

$$W_i^k, W_i^q \in \mathbb{R}^{d_{model} \times d_k}; W_i^v \in \mathbb{R}^{d_{model} \times d_v};$$

$$d_k = d_v = \frac{d_{model}}{heads}$$

3.4.2 Spatio-temporal encoding

Spatial Encoding :

Convolutional neural networks (CNN) are widely known for their outstanding performance on pattern recognition tasks, especially in computer vision tasks such as image classification. They receive as input raw images, and use a combination of filters and pooling layer to extract feature maps capable of detecting spatially-invariant patterns. To encode the spatial information of our input signal - a sequence of 8x8x1 frames, we applied a CNN at each position of the sequence (same CNN applied to all positions). The CNN architecture is depicted in Figure 3.3.

Positional Encoding :

In the multi-head self-attention mechanism information flows simultaneously. Consequently, there is no built-in sense of order like in recurrent neural networks (RNN's) where computation is interdependent : information from previous positions (time-steps) is used as input to the current position (time-step). In [139] authors proposed "Positional Encoding" in order for the model to incorporate sequence order. Positional encoding consists of adding a unique d-dimensional vector at each position defined as follow :

$$PE(pos, 2i) = \sin(pos/1000^{2i/d})$$

$$PE(pos, 2i + 1) = \cos(pos/1000^{2i/d})$$

where pos is the position and i is the depth ($i \in [1, d/2]$) and $d=d_{model}$ (figure 3.2).

Positional encoding encodes position in a similar fashion to binary encoding, excepts it uses floats instead of integers and "allow the model to easily learn to attend by relative positions, since for any fixed offset k, PE_{pos+k} can be represented as a linear function of PE_{pos} " [139]. In the example depicted in figure 3.1, adding positional encoding helps the model understand that the word "Thinking" comes before the word "Machines".

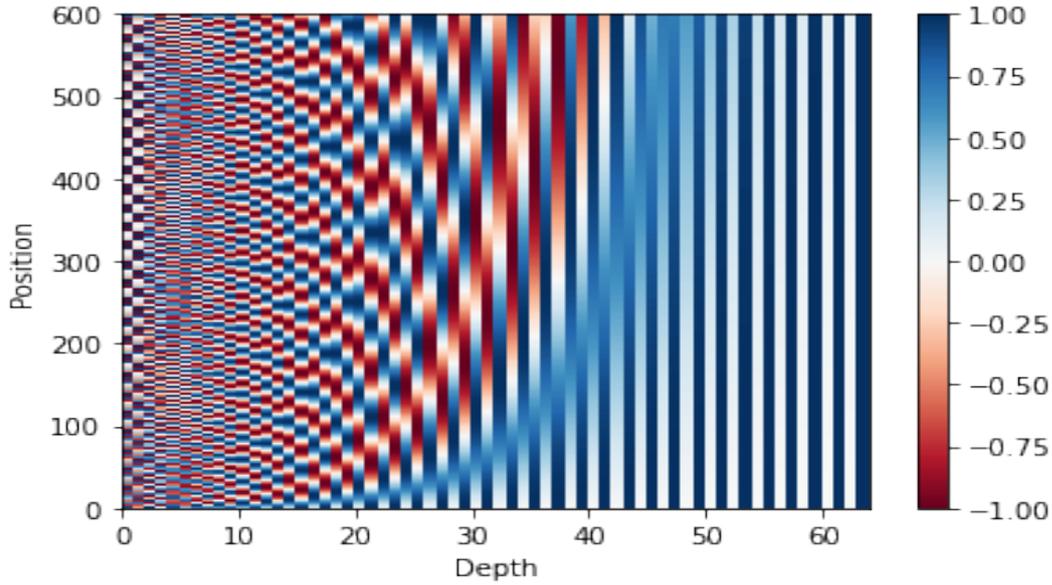


FIGURE 3.2 – Positional encoding used for CoST (sequence-length = 600; $d = 64$). For a position i we add a row i from this matrix to the CNN output vector ($d=64$) to encode position.

3.4.3 Inter-attention

We believe that not all frames contribute equally to the representation of touch gestures. By exploiting only a subset of the sequence, we allow the model to focus on relevant information thus enhancing classification. We introduce an attention mechanism based on the *Bahdanau attention* [8] and similar to the one used in the Sentence Attention in [148] :

$$score(h_t, qv) = v_a^T \text{Tanh}(W_1 h_t + W_2 qv)$$

$$\alpha_t = \frac{\exp(score(h_t, qv))}{\sum_i \exp(score(h_i, qv))}$$

$$context_vector = \sum_t \alpha_t h_t$$

where v_a^T is a fully-connected layer with 1 hidden unit that reduces the right term of the equation into a scalar score; qv (Query Vector) is the result of averaging all h_t vectors then applying a fully-connected layer (FC2) and $W_i \in \mathbb{R}^{d_{att} \times dim(h_t)}$. The importance of each h_t is determined by its similarity to the query vector (Figure 3.3). Put simply, inter-attention role is to select relevant information from the self-attention output sequence h_t .

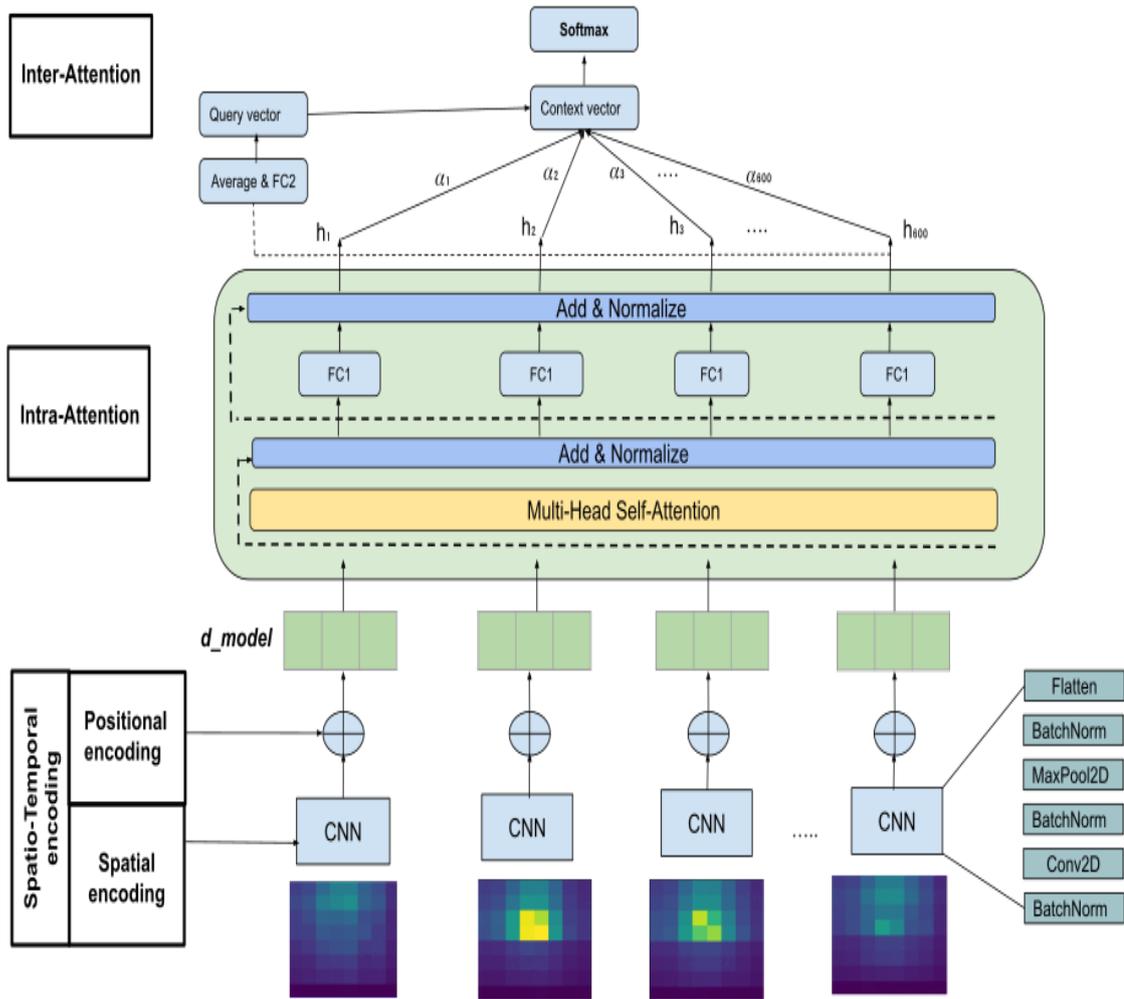


FIGURE 3.3 – Attentive Touch Model : architecture

3.5 Touch gesture recognition

3.5.1 Data Preprocessing

The CoST dataset contains variable length sequences while our model receives fixed length input. We noticed that 90% of data length is below 600 frames, therefore we chose to fix the length of touch gesture sequences to 600 to reduce the model complexity by using zero padding (8x8 zero matrices) for sequences length below 600 frames and truncation for sequences length above 600 frames. As for HAART, data length is already fixed to 432 frames.

For both datasets we tried global and channel-wise MinMax scaling and Standardization so that the model learn faster. Surprisingly, we have found that using a Batch-Normalization layer for normalizing the input data gave better accuracy scores.

3.5. Touch gesture recognition

TABLE 3.2 – Evaluation : Baselines and Attentive Touch Model

Dataset	Paper	Classifier	Accuracy
CoST	Ta et al.[133]	random forest	61.3%
	Gaus et al.[53]	random forest	58.7%
	Hughes et al.[71]	CNN-RNN	52.8%
	Hughes et al.[71]	CNN	42.3%
	Hughes et al.[72]	logistic regression	47.2%
	Balli Altugl et al.[11]	random forest	26.0%
	<i>Our method</i>	<i>ATM</i>	60.9%
HAART	Ta et al.[133]	random forest	70.9%
	Hughes et al.[72]	logistic regression	67.7%
	Gaus et al.[53]	random forest	66.5%
	Balli Altugl et al.[11]	random forest	61.0%
	Hughes et al.[71]	CNN-RNN	61.3%
	Hughes et al.[71]	CNN	56.1%
	<i>Our method</i>	<i>ATM</i>	67.8%

3.5.2 Hyperparameters Tuning and training

We performed a leave-one-subject Cross-validation on 21 subjects for CoST and 7 subjects for HAART (training sets) to determine the best hyper-parameters. At this stage the test set is left untouched. Our search space was composed of a Learning rate : ($7e-4$, $5e-4$, $4e-4$); Batch size : (16, 32, 64); Pool size, Filters and Kernel size : (2,3,4); FC1 and FC2 : (3, 12, 32); number of Heads : (8,16,32); epoch : (200); optimizer : (Adam [85]); loss-function : (categorical cross-entropy). Furthermore, we used Tensorflow [1] earlystopping callback with a patience of 30 epochs on validation-loss to reduce training time (Stop training when a validation-loss has stopped improving) and used Modelcheckpoint callback to save the weights of the model with the lowest validation-loss.

The search space contains 6561 combinations, therefore we opted for a random search over a grid search strategy [16]. We run the random search on a sample of 100 then picked the hyper-parameters that maximize the cross-validation accuracy score.

The best hyper-parameters we found are in Table 3.2. Also, the best batch size is 32, best learning rate is $4e-4$ and the best epoch is 200. We trained the model on the whole training sets for 200 epochs using the aforementioned hyper-parameters then classified the test sets according to the social touch challenge protocol (10 subjects for CoST and 3 subjects for HAART).

3.6 Results and Discussion

The classification results were reported in Figure 3.5" comprised two confusion matrices. Looking into HAART confusion matrix we notice that the "constant", "no-touch", "scratch" and "stroke" touch gestures were mainly classified correctly while "tickle" and "rub" gave the lowest accuracy scores. Furthermore, "tickle" and "rub" were confused with "scratch", and "pat" was confused with "stroke". For CoST, the confusion was between (hit, slap), (grab, squeeze), (pat, slap), (scratch, tickle), (stroke, rub) and (tap, pat). These findings are coherent with previous studies [81] in which the mis-classification was attributed to several factors such as : similarity of touch gestures, variations (gentle squeeze is similar to a normal grab) and inter-subject variability.

We have obtained a test accuracy score of 61% for CoST and 68% for HAART. Our results are comparable to the studies found in Table 3.2. Other studies achieved higher accuracy scores [3] [80] [146] but we only kept the ones that followed the social touch challenge protocol. Even though our model does not outperform the Ta et al. model [133], we have achieved similar results with a deep learning model on both datasets without resorting to feature engineering.

Furthermore, our model inherits the advantages of the Transformer model [139] such as : (i) modeling longer sequences : self-attention models the relationship between all positions non-sequentially. RNN's reads one position at a time thus information from the position 0 will barely reach the position 600; (ii) gaining insight into how information flows by visualizing self-attention weights. The limitation of this approach is the size of data, deep learning models require larger datasets to perform well which is difficult to achieve in the context of social touch recognition.

3.6. Results and Discussion

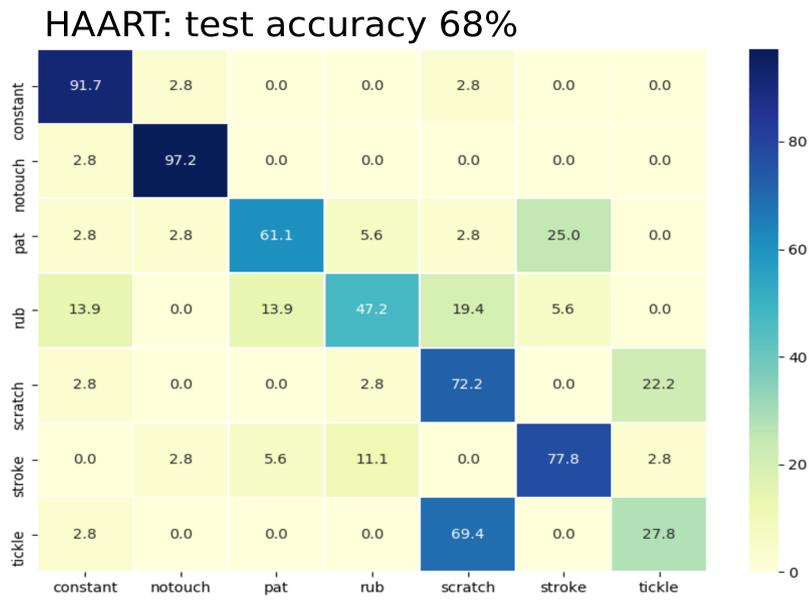


FIGURE 3.4 – Confusion matrix for HAART dataset

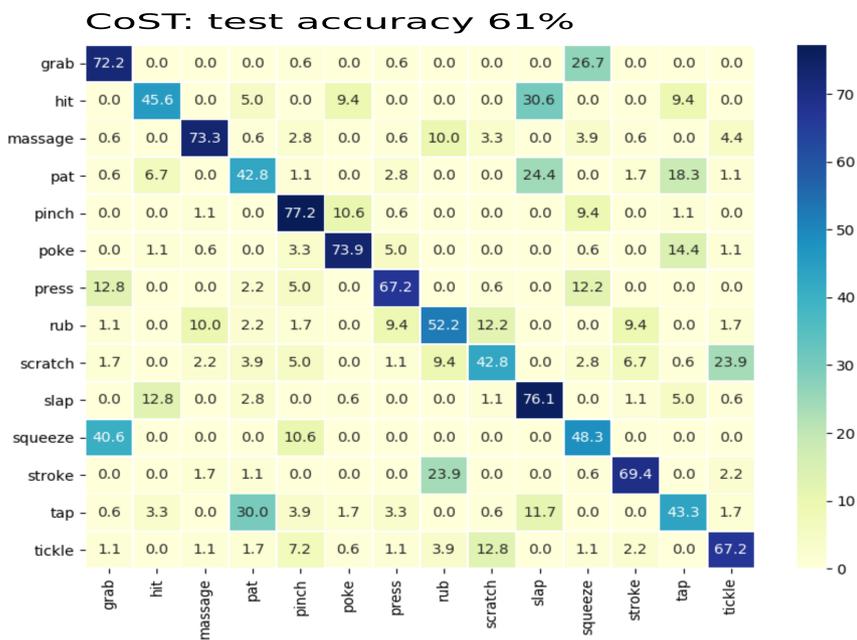


FIGURE 3.5 – Confusion matrix for CoST dataset

3.7 Conclusion

This chapter presents an attention-based neural network architecture for touch gesture recognition. Our model performed well on both datasets : 61% on CoST and 68% on HAART. Although it did not outperform the best scores (61.3% on CoST and 70.9% on HAART) this architecture does not require feature engineering, and uses self-attention as an alternative to RNN's. Furthermore, we believe that the performance of this architecture scales with the size of datasets therefore building generative models for data augmentation is the next step for enhancing the performance of our social touch recognition systems.

Chapter 4

Stimuli-Imitation : Elicitation Approach for Social Touch Gesture Data Collection

4.1 Introduction

In affective computing, two approaches [107] are usually considered to elicit emotions : **subject-elicited** and **event-elicited**. In the first one, subjects are asked to remember or to act an emotional episode. In the second, an emotional stimuli is selected and used to elicit the desired valence and arousal levels. The most prominent event-elicited methods used : emotional images [18]; emotional film clips [118]; music and ideation [38].

One of the main issues with touch data collection is the elicitation of socio-affective touch gestures. Unlike other modalities such as speech or face, subjects have to effectively touch sensors, therefore most approaches for touch elicitation are acted (**subject-elicited**). Furthermore, in the literature (Chapter 2 : Touch Elicitation Approaches section), we found that subjects relied solely on their individual interpretations and experience to convey an emotion or a social message through touch. In other words, when provided with emotional labels (e.g Anger), subjects were generally instructed to think about a plausible way to communicate these emotions to a touch interface. Here, factors that modulate the interpretation and the expression of touch (Figure 2.1) were not controlled nor incorporated into the experimental design, thus subjects touch expressions were not constrained by these factors. As example, a subject given the label "Anger" could recall a moment when He/She was angry at a spouse for cheating; while another subject could remember a situation where He/She was angry at His/Her child for not doing homework. Though both situations are attributed to the same label, their touch expression are expected

to be highly different. Consequently, touch data collected this way is context-free, socially-divorced and exhibit high inter-subject variability [81].

When developing a touch recognition system, high inter-subject variability could be harnessed to improve generalizability, by collecting more data and adding more sensors. However, in the case of Human-Machine Interaction, collecting lots of data is time consuming (1000 samples=1000 interactions) and adding more sensors is expensive, specially in a full-body touch interaction.

Furthermore, a touch recognition system built upon a label-based data collection approach, will be capable of recognizing touch gestures (e.g Hug) and Emotions (e.g love) separately, but will fail to capture the synergy between them. As an illustration, a loving hug could be expressed differently depending on the circumstances : (i) hugging a friend that you haven't seen in many years ; (ii) hugging a spouse before leaving for a long trip. For a label-based system the predicted labels for both scenarios are "Hug" and "Love". Though these scenarios differ in meaning and expression, they are considered identical by the system.

Another prevalent problem in affect detection systems is the Experience vs Expression of affect [22]. A subject expressing an emotion does not necessarily imply that She/He is experiencing it. Similarly, a system that could predict "Anger" from a touch could not guarantee that the subject touching the device is at an angry state, nor approximate the real emotional state. To tighten the gap between the experience and the expression of affect, instead of using abstract labels, an elicitation procedure should provide subjects with real situation that evokes an emotional response.

In chapter 2 (Table 2.1) we presented different studies that used a label-based approach for data collection. In the next section, we will present a novel way of collecting touch data that overcomes the aforementioned shortcomings.

4.2 Stimuli-imitation framework

In this section, we propose to explore a new approach that aims to improve social touch data collection that we named **stimuli-imitation**. The key idea is to ask participants to imitate a social interaction stimuli where social touch is employed by the partners. The social interaction stimuli are extracted from the Socio-Affective Touch Database (SATD) [89], which is composed of videos of non-professional actors that have been annotated in both valence and arousal. Actors were given scenarios that evoke a certain touch communication in order to perform, therefore the database is a mapping between scenarios and their implementations by actors (videos).

In our stimuli-imitation protocol (Figure 4.1), we ask subjects to : (i) read a scenario (Table 4.1) and watch a video related to the scenario (Figure 4.3), (ii) rate stimulus

using the Self-assessment Manikin [17], and (iii) imitate the observed socio-affective touch on a mannequin equipped with pressure sensors (Figure 4.7). The protocol is composed of two steps : the Socio-emotional encoding and Imitation.

4.2.1 Socio-emotional encoding

The socio-emotional encoding step aims at eliciting context-specific emotions. In stimuli-imitation, an elicited emotion is modulated by reading a scenario, and induced by observing a social touch video varying in valence and arousal (event-elicited). Social touch can induce feelings in an observer who is not part of the touch itself [105], and adding a scenario provides contextual information (e.g social context, relationship) ensuring that participants could disambiguate between similar touch gesture (e.g hug) that holds different meanings (e.g consolation or love). Afterwards, participants annotated in terms of valence and arousal our stimuli consisting of the combination of a scenario and a video.

4.2.2 Imitation

After the emotion elicitation step, participants were asked to imitate the observed social touch – a complex dynamic visual input– into a touch gesture. The observed social touch video duration is 3 seconds, long enough to guess the purpose and the affective meaning, but not sufficient enough to replicate exactly the same touch gesture. Imitation is goal-directed [15], subjects observes and replicates the goal of other’s behaviors or actions. Imitation is different from mimicry : subjects observes and only replicates another’s behaviors or actions without the aim to achieve the same goal. Imitation is a conscious act, not every part of an observation sequence is necessarily replicated. We argue that imitation of another’s socio-affective touch gesture will allow to constrain the elicitation procedure resulting in collecting prototypical and representative touch gestures.

4.2.3 Framework Evaluation

As we discussed earlier, by employing labels to elicit socio-affective touch gestures we collect context-free and socially-divorced data. We believe that transitioning to a scenario-based elicitation approach could enable us to include social context therefore model a wider range of socio-affective touch interactions. In their study [96], McIntyre et al. showed that close adults relationships could communicate affect through touch using scenarios as stimuli, for instance : *"Your partner has just told a funny joke. Try to communicate AMUSEMENT through touch"*.

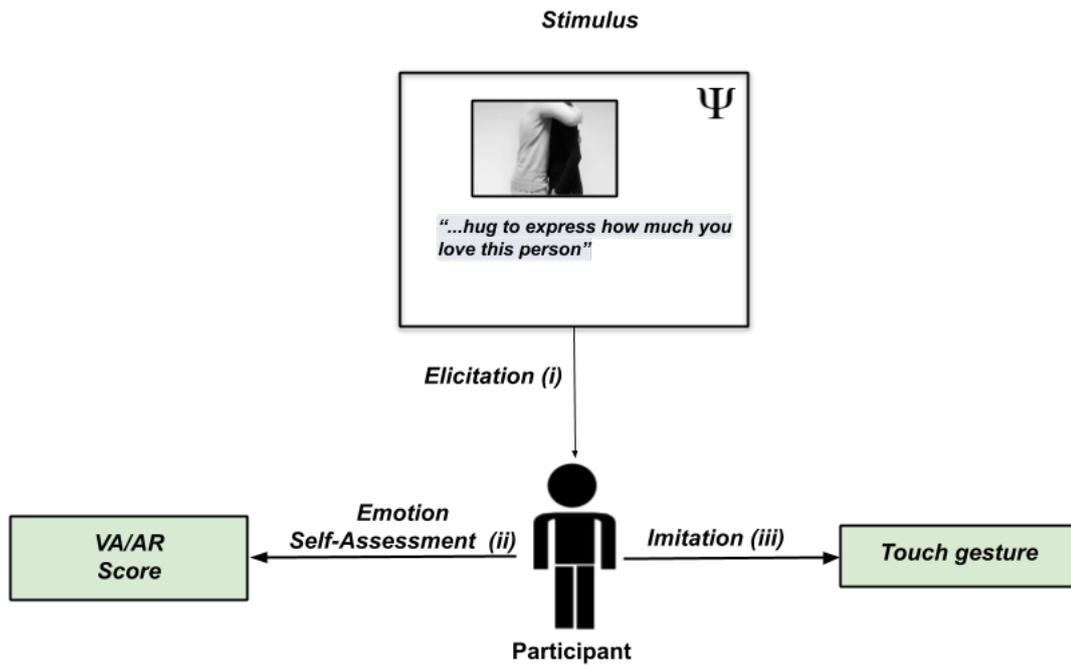


FIGURE 4.1 – Stimuli-imitation protocol : (i) participant read a scenario and watch the associated video (stimulus) taken from SATED [89]; (ii) Evaluate in term of valence and arousal the stimulus; (iii) imitate the observed touch gesture on the mannequin.

In the context of Human-Machine Interaction, we propose to evaluate the Scenario condition (S) (Figure 4.2) which consists in only providing a scenario to the participant. The evaluation is done by training and testing a machine learning model on touch data collected using the scenario stimulus from SATED. Though the scenario stimulus provides participants with social context, it does not prevent the high inter-subject variability problem in that elicited touch gesture varies between participants even for the same situation. Therefore we propose to compare the Stimuli-Imitation (SI) condition with the Scenario condition (S), the Stimuli-Imitation is the Scenario condition enriched with the video experience and replication. The aim here is to evaluate the efficiency of Stimuli-Imitation elicitation framework to produce consistent socio-affective touch gestures. For this purpose, we propose to :

- Compare the two conditions (S and SI) in terms of Touch gesture and Socio-affective Touch gesture (SATg) classification (see Table 4.1).
- Cross-compare Machine Learning models trained respectively on S and SI data to find out which condition generalize better.

In the SATED database, participants watched then annotated the videos in terms of

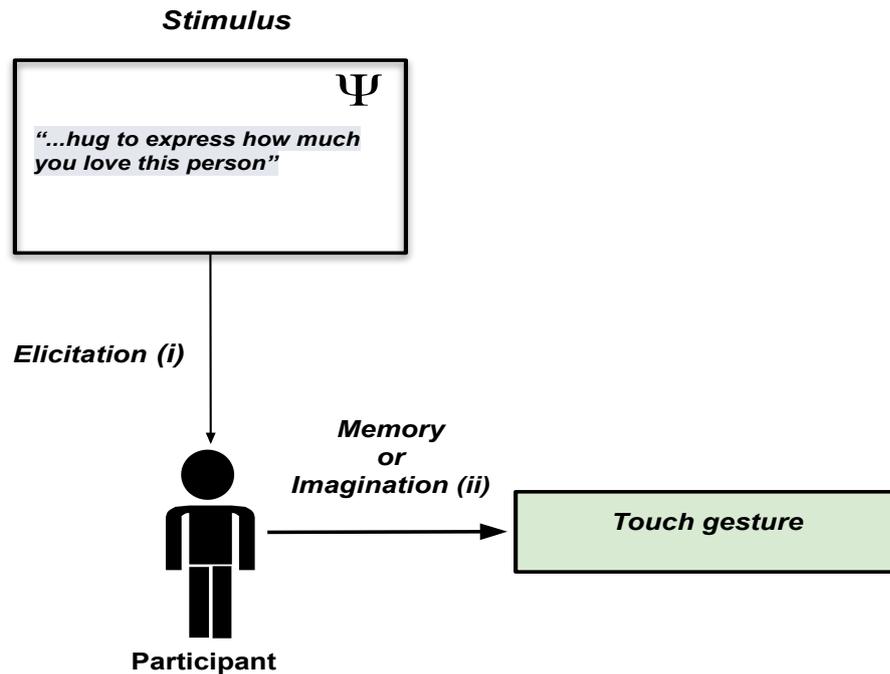


FIGURE 4.2 – Scenario protocol : (i) participant read scenario taken from SATED [89]; (ii) participant recall from past experience or use his imagination to produce the touch gesture.

valence and arousal. In Stimuli-imitation (SI), beside watching the videos, participants had also access to the scenarios. Consequently, we propose to compare the affective experience induced into participants for Stimuli-only condition (video only) and Stimuli-Imitation condition (video and scenario) to measure the effect of introducing a scenario on the annotations. Here we compare :

- Intra- and Inter-subject annotations consistency.
- Valence and Arousal annotations.

Finally, we investigate the influence of individual traits : empathy, personality on the Stimuli-imitation (SI) valence and arousal annotations. In Stimuli-imitation we used videos of social touch interaction to elicit affective response in participants. A study done by Peled-Avron et al. [105] showed that the perception of social touch is modulated by empathy. Similarly, another study suggests that the brain vicarious response to observed touch interact with participant personality traits [119]. We believe that social touch is experienced differently from a person to another, and that SI stimuli are perceived differently depending on empathy and personality scores of participants.

4.3 User study

In our study, participants will interact tactually with a mannequin equipped with pressure sensors (details in the subsequent section). The interaction involves the upper-body of both mannequin and participants. To collect useful data, sensors should be placed on the mannequin in accordance with the stimuli. In the Stimuli-imitation condition (SI), participants imitate the touch interaction displayed in the videos, therefore sensors should be placed in accordance with the videos. Though we ignore where the Scenario condition (S) participants will touch the mannequin, if their touch expression will not deviate highly from the SI condition in that the videos are an implementation of the scenarios. A first step is to select suitable stimuli from SATED for our study.

4.3.1 The Mannequin

The mannequin (Figure 4.7) was composed of a fiberglass bust covered with a linen fabric tied to an adjustable stand. To ensure the stability of the overall structure and to prevent the mannequin from tilting to the left or to the right, the height of the mannequin was fixed at 175 cm (average male height in France), and two heavy metallic bars were screwed to the end of the stand. Two cylindrical arms were attached to the bust : (1) a left arm made of cardboard, and (2) a 3D printed arm made of plastic that serves as support for symmetrical touch gestures that requires touching both arms (e.g shake).

4.3.2 Stimuli selection & Sensors placement

Stimuli selection : the SATED (Socio-affective touch expression database) [89] comprised 13 stimuli (Figure 4.3) dedicated to social touch between humans, and 12 stimuli where humans touch objects. The later were not considered in this study. From the 13 stimuli we excluded the ones that : (1) Are not feasible on a mannequin, (2) Both actors initiate touch gestures at the same time, (3) Involves the mannequin hand thus require a hand-sensor. Accordingly, we were left with the following stimuli :

- 4 Positive touch gestures : 2 hugs (hug_2p, hug_3p) and 2 strokes (str_1p, str_2p).
- 3 Negative touch gestures : 2 shakes (sha_1n, sha_2n) and 1 slap (slap_1n)

Initially we had selected also a neutral Touch gesture : Tap1_neu. Even though it respected the aforementioned selection criterion, it was removed because no signal was detected in the pilot study.

Sensors placement : The selected stimuli were given to six annotators (colleagues from the lab) that watched the stimuli videos, then indicated where the touch occurred on both actors body (Male and Female) using the form depicted in Figure 4.4.

4.3. User study

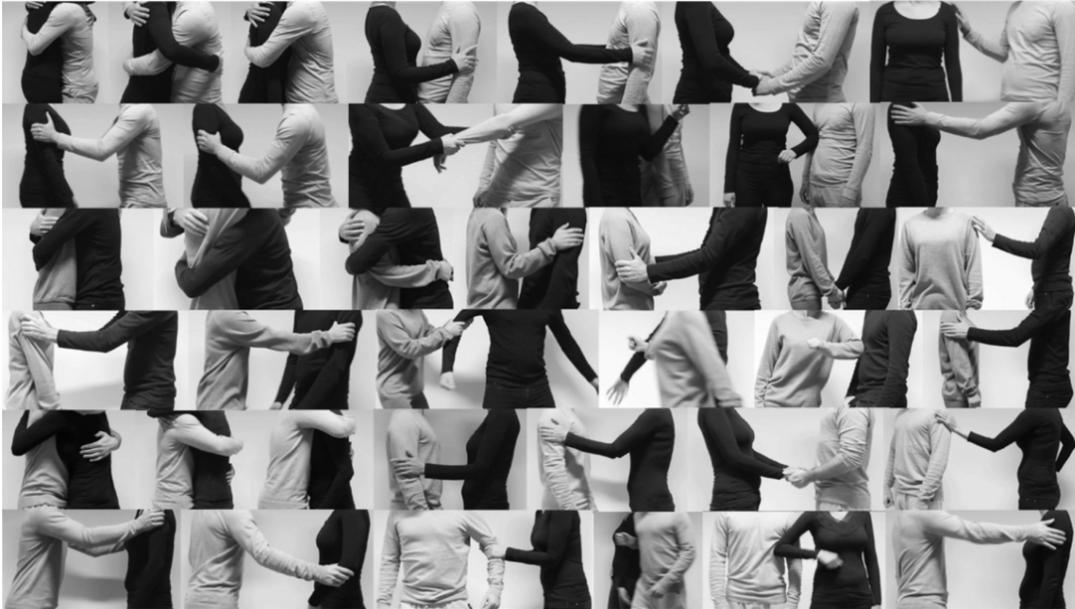


FIGURE 4.3 – This figure shows 13 stimuli taken from SATED [89]. Each stimuli is composed of 3 socio-affective touch expression videos varying in valence and arousal.

Number	Body Part
1	Forearm
2	Arm
3	Shoulder
4	Chest
5	Abdomen
6	Upper-Back
7	Lower-Back

	Hand	Forearm	Arm	Shoulder	Chest	Abs	upper back	lower back
Male	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Female	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FIGURE 4.4 – The figure shows the annotation form and the body-part map used by annotators

The purpose of the annotation session is to figure out where to place sensors on the mannequin in order to collect relevant data thus we only kept the touch receiver annotations. The results summarized in Figure 4.5 show that by placing a sensor on the *Arm* and another one on the *Shoulder* we could cover all the stimuli.

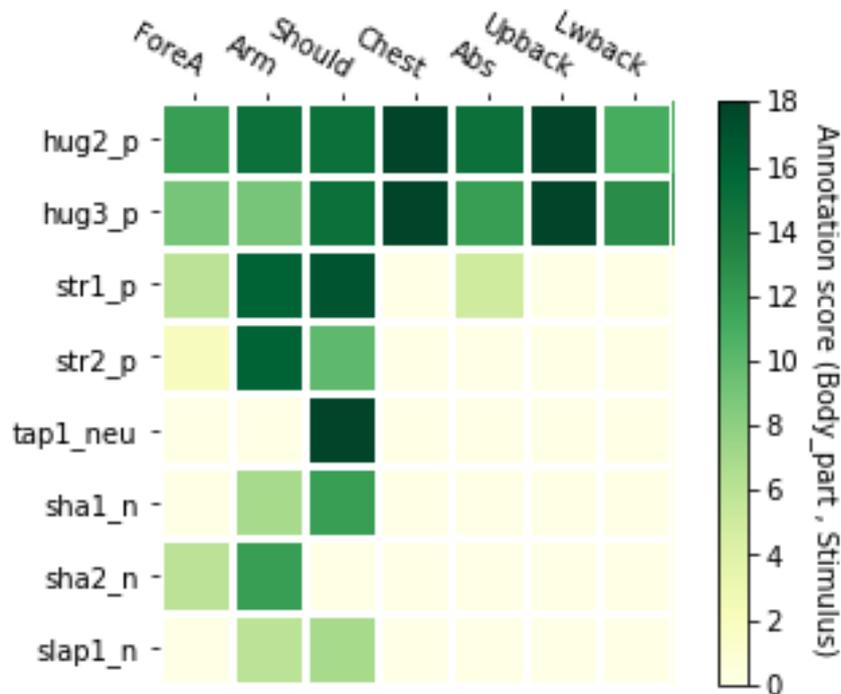


FIGURE 4.5 – For each selected stimulus (3 videos) six annotators indicated where the touch-recipient was touched. For instance, in the Tap1_neu stimulus the shoulder was the only body-part touched.

From the annotation data, we found that some body parts were more frequently touched than others (from the touch receiver perspective), therefore covering the mannequin full-body with sensors is not necessary. We decided to place two sensors on : left-arm and left-shoulder. However to discriminate between hug_2p and hug_3p; and between a hug and a stroke for instance, we need to place more sensors on the mannequin. Consequently, we added a sensor on the *Chest* and another one on the *Back* (covering 50% upper-back and 50% lower-back). Furthermore, no sensor was placed on the *Forearm* because no signal was detected in pilot study.

For sensing touch gesture, we used the *plug-and-wear* 12x12 HIGHDYN pressure sensors (Figure 4.7). Each sensor recorded a 12*12 pressure matrix (144 channels) at a sampling rate of 30 Hz, and each value channel ranged between 0 to 255. Consequently, the mannequin sensors collects pressure and position data simultaneously from the aforementioned four body parts : left-Arm, left-Shoulder, Chest and back. To hide sensors from participants, we dressed the mannequin with a soft fabric track-suit jacket (Figure 4.6).

4.3. User study

Each stimulus consists of 3 videos, to reduce the length of the experiment we selected only 2 videos for each stimulus (Table 4.1)

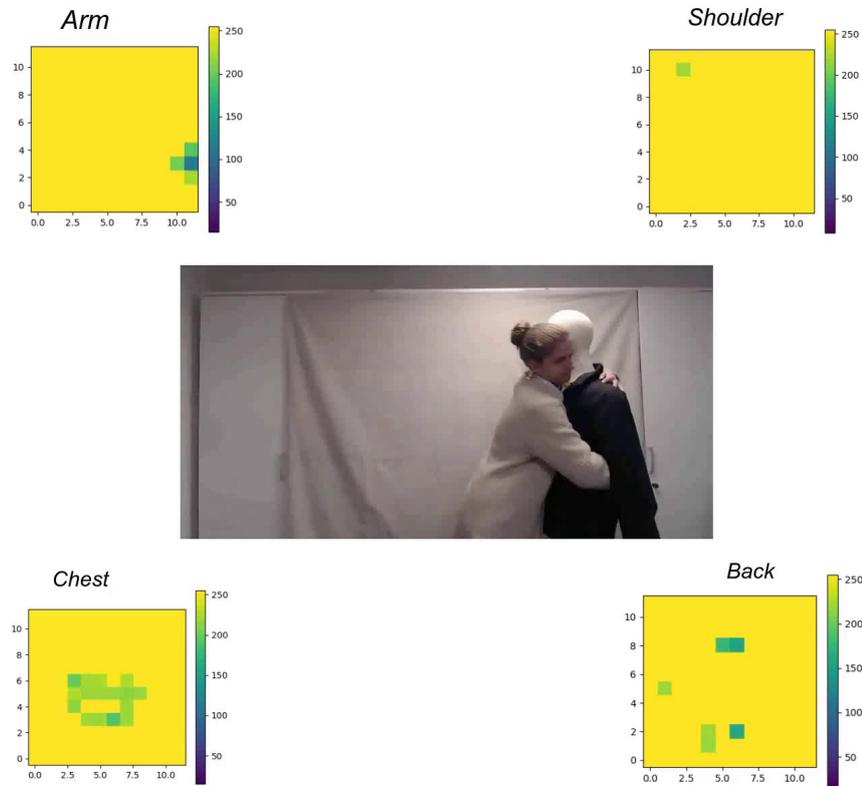


FIGURE 4.6 – Participant hugging the mannequin.

TABLE 4.1 – Selected stimuli.

Touch gesture	SAT gesture	Scenario	Videos
Hug	hug2_p	Your partner looks somehow lovelier today. You want to give a hug to express how much you love this person.	15, 28
	hug3_p	Your sibling just ended the long-term relationship. You want to console him (her) by hugging	16, 29
Stroke	Str1_p	You want to flirt with him (her) by stroking his (her) arm with intimacy.	17, 30
	Str2_p	Your friend is crying. You want to console this person by stroking on his (her) arms.	18,31
Shake	Shake1_n	You just found out that your partner cheated on you. You are very disappointed and angry. You ask him (her) how and why this happened aggressively while shaking his (her) whole body.	21, 34
	Shake2_n	Your sibling committed crime. You are sad and angry since you cannot understand why he (she) made such a horrible decision. You firmly ask him (her) how and why this has happened while shaking his (her) arm.	9, 22
Slap	Slap1_n*	Your sibling always makes the same mistakes. You slap his (her) arm to make him (her) realize that you are annoyed and that you expect something better.	13, 26



FIGURE 4.7 – The matrix sensor (right side) render position and pressure data when touched. Four matrix sensors are placed on the mannequin left arm, chest, left shoulder and back. The purpose of the mannequin-sensor is to collect social touch data from various regions of the body simultaneously.

4.3.3 Study setup

The study was conducted in two separate rooms at the INSEAD Behavioral Lab : (i) a room where participants filled two questionnaires (described in Individual traits assessment section), (ii) and another one where the experiment was held. The protocol was approved by the “INSEAD Behavioral Lab” ethics committee.

In the experimental room depicted in (Figure 4.8) a desktop computer was connected to the mannequin-sensor and to a camera. On the other side of the room, a wireless keyboard was placed on a table within the participant arm’s reach. The distance between the tape and the mannequin-sensor is approximately 120 cm; and distance between the tape and the camera is approximately 190 cm. For the display of stimuli, a dell screen (52cm x 30cm) was placed at approximately $\alpha=30^\circ$ to right of the participant at also 120 cm. All the devices were orchestrated by a PsychoPy script.

4.3.4 Participants

40 right handed participants (21 females) aged between 19 and 31 (Mean=23.2, SD=3.04), with a height between 148 cm and 190 cm (Mean=169.8, SD=8.88) were recruited by the INSEAD behavioral lab. Prior to the experiment, each participant signed a consent form and was assigned randomly to one of the two conditions (20 participants for each condition). Afterwards, the participant was led to a room where the instructions were explained.

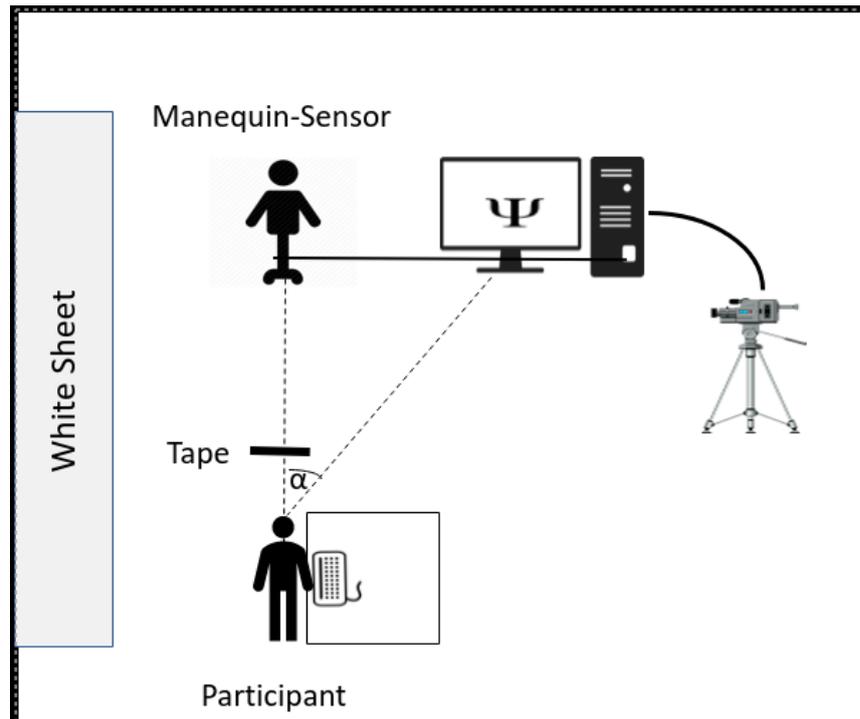


FIGURE 4.8 – Experiment setup

4.3.5 Individual traits assesment

To asses the influence of dispositional empathy and personality on the perception of the SI stimuli (video and scenario), participants were asked to fill, at the end of the experiment, the following questionnaires :

Interpersonal Reactivity Index (IRI)

IRI is a self-report method for measuring dispositional empathy [34]. In contrast to situational empathy which is a response to a specific eliciting situation, dispositional empathy is seen as a stable character trait. It measures people inclination toward adopting others feelings and point of views. It consists in 28-items answered on a 5 point likert scale ranging from “Does not describe me well” to “Describes me very well” and equally divided among four distinct subscales :

- Perspective taking : the tendency to spontaneously adopt the psychological view of others in everyday life
- Empathic concern : the tendency to experience feelings of sympathy or compassion for unfortunate others
- Personal distress : tendency to experience distress or discomfort in response to extreme distress in others
- Fantasy : the tendency to imaginatively transpose oneself into fictional situations

Big Five (BF)

The big five framework is a taxonomy based system for personality traits measurement. It is widely used by researchers to explore relationships between individual differences in personality and a variables of interest such as mental health or academic performance. It defines personality in terms of five core factors that could be understood through the following traits :

- Openness : Imagination, Creativity, Curiosity (10 items).
- Conscientiousness : Persistence, Reliability (9 items).
- Extraversion : Sociableness, Warmth, Energy (8 items).
- Agreeableness : Altruism, Trust, Loyalty (10 items).
- Neuroticism : Awkwardness, Nervousness, Anxiety (8 items).

In this study we used a french version called BFI-Fr [108]. The questionnaire contains 44 items not equally divided among the five factors.

4.3.6 Participants self-assessment of Valence and Arousal

Emotions were assessed using the SAM (self-assessment manikin). SAM is a non-verbal pictorial technique that measures the affect experienced while processing an emotional stimuli according to valence, arousal and dominance (not measured here) using a likert scale (e.g 1-9). A sheet of paper containing both the valence and arousal SAM was placed on the table next to the wireless keyboard (input device). At each end of the SAM (valence and arousal) affective states were given for reference to help participants (Figure 4.9).

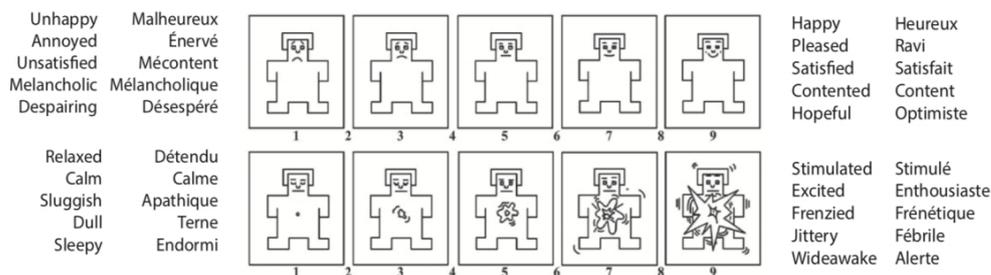


FIGURE 4.9 – Self-assessment Manikin [17]

4.3.7 Experiment protocol

The experiment included three successive parts : (i) familiarization with the mannequin-sensor, (ii) a condition-dependent task (main task) repeated three times, (iii) two questionnaires filled in a separate room.

First, participants were introduced to the mannequin-sensor and were asked to touch every sensor in this order : left-arm, left-shoulder, chest and back while watching on the screen the evolution of the output signal (12*12 pressure time series) for five minutes. At this stage, participants are aware of the exact location and responsiveness of the sensors placed on the mannequin. Next, participants were asked to perform a trial of the main task. For the main task, two conditions are considered :

Stimuli-imitation (SI) condition is composed of two successive tasks (Figure 4.10) : (i) Elicitation & Evaluation ; (ii) Imitation. In the Elicitation & Evaluation task, a scenario portraying a social situation centered around touch (sated scenario) and its implementation (associated sated videos) were displayed (Table 4.1). Participants were asked first, to read carefully the scenario, then locate in the video the person -identifiable by shirt color- that initiate the touch gesture. Then, while identifying with the touch initiator, to assess their feeling about the communicated socio-affective message, by considering the contextual information from the scenario and the visual cues provided by the sated video. The assessment of emotions was done using the Self-Assessment Manikin (Figure 4.9) tool on a scale of 1-9 for valence and arousal.

In the imitation task, participants were asked, similarly, to identify with the actor that initiate the touch gesture, then imitate on the mannequin-sensor the observed touch gesture with the same energy level. An exception was made for the *stroke* touch gesture. The fact that we are using pressure sensors makes it difficult to detect light strokes, therefore we instructed participants to apply a reasonable force when stroking : sufficient to activate the sensor (familiarization phase helped here) without hurting the spontaneity of the touch gesture. A set of instructions were given to guide participants through the data collection phase of touch gestures :

1. Stand in front of the mannequin-sensor while putting both feet on the tape (figure 4.8).
2. Press space on the wireless keyboard to start the recording process.
3. Imitate the touch gesture on the mannequin-sensor after hearing the first beep sound.
4. Stop when hearing the second beep sound (interval between sound beeps is 8 seconds).

This process is repeated for all the selected stimuli.

4.4. Social Touch Recognition



FIGURE 4.10 – Stimuli-imitation tasks.

Scenario (S) condition contains only the Scenario task (Figure 4.11). Participants were given a sated scenario and a touch gesture label (e.g hug) to apply on the mannequin-sensor. Unlike stimuli-imitation, participants were asked to apply the touch gesture as they see fit. The only constraint was to touch at least one body-part equipped with a sensor. Same instructions as stimuli-imitation were given to guide participants through the data collection phase except for the imitation part. Here, participants were instructed to use their imagination or memory to generate the touch gesture.

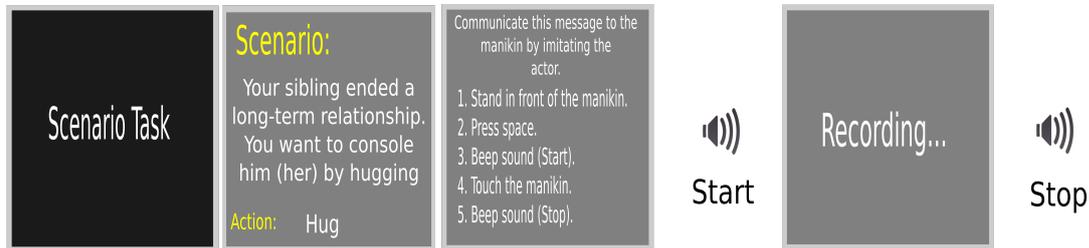


FIGURE 4.11 – Scenario task.

4.4 Social Touch Recognition

In this section we study the influence of visual cues and imitation on automatic social touch recognition by considering two conditions (Figure 4.12) :

- **SI** condition : Participants were provided with a scenario and a social touch interaction video, then were asked to imitate the touch interaction displayed, on the mannequin-sensor (Figure 4.7).
- **S** condition : Participants were provided only with a scenario (Table 4.1) then were asked to perform a touch gesture on the mannequin-sensor.

For Social Touch Recognition we consider two tasks : *Touch gesture* recognition and *Socio-affective Touch gesture* (SATg) recognition. In the former, the purpose is to classify 4 touch gestures (Hug, Stroke, Shake, Slap) regardless of the scenarios in which they are implemented, while in the latter we classify 7 touch gestures (hug2_p, hug3_p, Str1_p, Str2_p, Shake1_n, Shake2_n, Slap1_n) that holds a specific socio-affective meaning. Its important to note that, for instance, Str2_p is a special case of Hug (hierarchical relationship).

Machine learning models were trained and tested for both tasks using data collected under the two conditions (SI and S), and performance was measured using *accuracy* and *F1* scores. To determine which data collection method generalizes well we cross-compared the models, by training on SI data then testing on S, and vice versa.

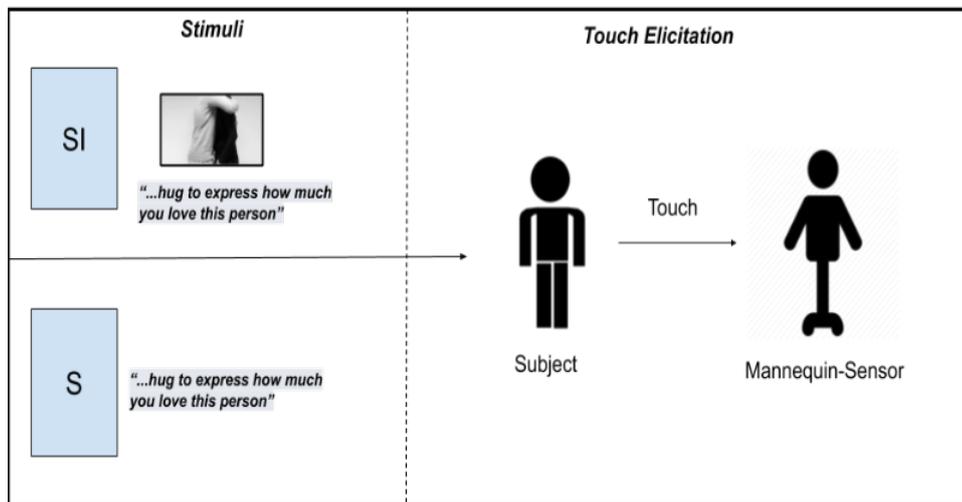


FIGURE 4.12 – Difference between an SI and S stimulus in the case of a hug touch gesture.

4.4.1 Train and Test set

Data was collected from 40 participants, 20 for each condition :

In the **Stimuli-imitation (SI)** condition, 840 samples (20 subjects x 14 videos x 3 repetitions) were divided into a training set (70%, 588 samples, 14 subjects) and a test set (30%,252 samples, 6 subjects).

In the **Scenario (S)** condition, 840 samples (20 subjects x 7 scenarios x 6 repetitions) were divided similarly into a training set (70%, 588 samples, 14 subjects) and a test set (30%,252 samples, 6 subjects). The training/test division was done randomly on subjects for both conditions.

The Slap touch gesture contains only one SATg : Slap1_n (Table 4.1) therefore the dataset is imbalanced. To balance the dataset we synthesized new examples for the Slap

touch gesture using Synthetic Minority Oversampling Technique or SMOTE [26]. SMOTE was applied on the features described in the next section. The final number of samples for each condition is therefore 960 instead of 840. Adding a synthesized Slap1_n is equivalent of adding two extra videos for SI, or adding one scenario for S.

4.4.2 Feature Extraction

In the chapter 3, we found that the literature on social touch gesture recognition divided features into three groups [134] [146] : (1) Global features that represent overall statistics of the gestures ; (2) Channel-based features which focus on the relationship between channels ; (3) Sequence features that describe the change of the gesture. Then these features were combined and used as input to a model. In our case we used only Global features and Sequence features defined as follow :

Global features :

- Average pressure on the 144 channels over all frames (1 feature);
- Maximum value of pressure of all channels over all frames (1 feature);
- Standard deviation (SD) of pressure of all channels over all frames (1 feature);
- Mean and SD of pressure over all frames of each column (24 features);
- Mean and SD of pressure over all frames of each row (24 features);
- Mean and SD of pressure over all frames of each diagonal (24 features);
- Mean and SD of pressure over all frames of each opposite-diagonal (24 features);

Sequence features :

- Fast Fourier transform based features : these features are the results of applying the Fast Fourier operator to the sequence of average pressure. The first 16 highest frequencies values are used as features (16 features);
- Discrete cosine transform based features : these features are the results of applying the DCT (type 2) operator to the sequence of average pressure. The first 20 values are used as features (20 features).

We extracted 135 features for each sensor. Therefore the total set of features used as input for our model is 540 (135 features x 4 sensors).

4.4.3 Model and Training

A Support Vector Machine (SVM) was used as a baseline for Touch gesture and Socio-affective touch gesture (SATg) recognition. First we normalized the training/test

data for each condition separately by removing the mean and dividing by the standard deviation. Then, a *GridSearch* from the *Scikit-learn* library was used to find the hyper-parameters that maximize the cross-validation score (3-fold cross-validation) on the training data. The hyper-parameters search space was composed of a Kernel : ['Linear', 'rbf', 'poly'], a regularization parameters C : [0.01,0.1,1,10,100,1000] and γ : [1,0.1,0.001,0.0001].

4.4.4 Results

The train/test split was repeated 30 times, and each time subjects were assigned randomly between the train (14 subjects) and test set (6 subjects).

Touch gesture and SATg recognition

The SVM models trained and evaluated on the Stimuli-imitation condition gave higher accuracy scores for Touch gesture recognition (81% against 67%) and Socio-Affective Touch gesture (SATg) recognition (60% against 43%) than the models trained on the Scenario condition (Figure 4.13).

In the case of Socio-affective Touch gesture recognition and in both conditions, an F1-score was calculated on binary confusion matrices composed of SATg (e.g str1_p and str2_p) that belong to the same Touch gesture (e.g Stroke) :

- Hug (SI-F1=0.65, S-F1=0.36);
- Stroke (SI-F1=0.75, S-F1=0.51);
- Shake (SI-F1=0.76, S-F1=0.70).

The Slap touch gesture contains only one SATg therefore it was excluded from this analysis. The model trained on stimuli-imitation was able to differentiate between SATg that belong to the same touch gesture more accurately than the one trained on Scenario condition.

Cross-Comparing Models

The SVM_S and SVM_{SI} models were trained respectively on the S and SI training sets. The Table 4.2 and 4.3 summarize the Accuracy results (Mean and SD) obtained on the S and SI test sets by the two models. We found that : (i) models trained on SI generalize well on S ($SVM_{SI}(S) \approx SVM_S(S)$) for both tasks; and (ii) models trained on S gave better results on SI than S ($SVM_S(SI) > SVM_S(S)$) for Touch gesture recognition. However, for SATg recognition we found that the gap is smaller ($SVM_S(SI) \approx SVM_S(S)$).

4.4. Social Touch Recognition

		Model	
		SVM_{SI}	SVM_S
Data	SI	0.81 (0.03)	0.74 (0.04)
	S	0.66 (0.05)	0.67 (0.04)

TABLE 4.2 – Cross-Comparing Touch Gesture models using accuracy scores.

		Model	
		SVM_{SI}	SVM_S
Data	SI	0.60 (0.02)	0.47 (0.02)
	S	0.40 (0.03)	0.43 (0.03)

TABLE 4.3 – Cross-Comparing Socio-Affective Touch Gesture models using accuracy scores.

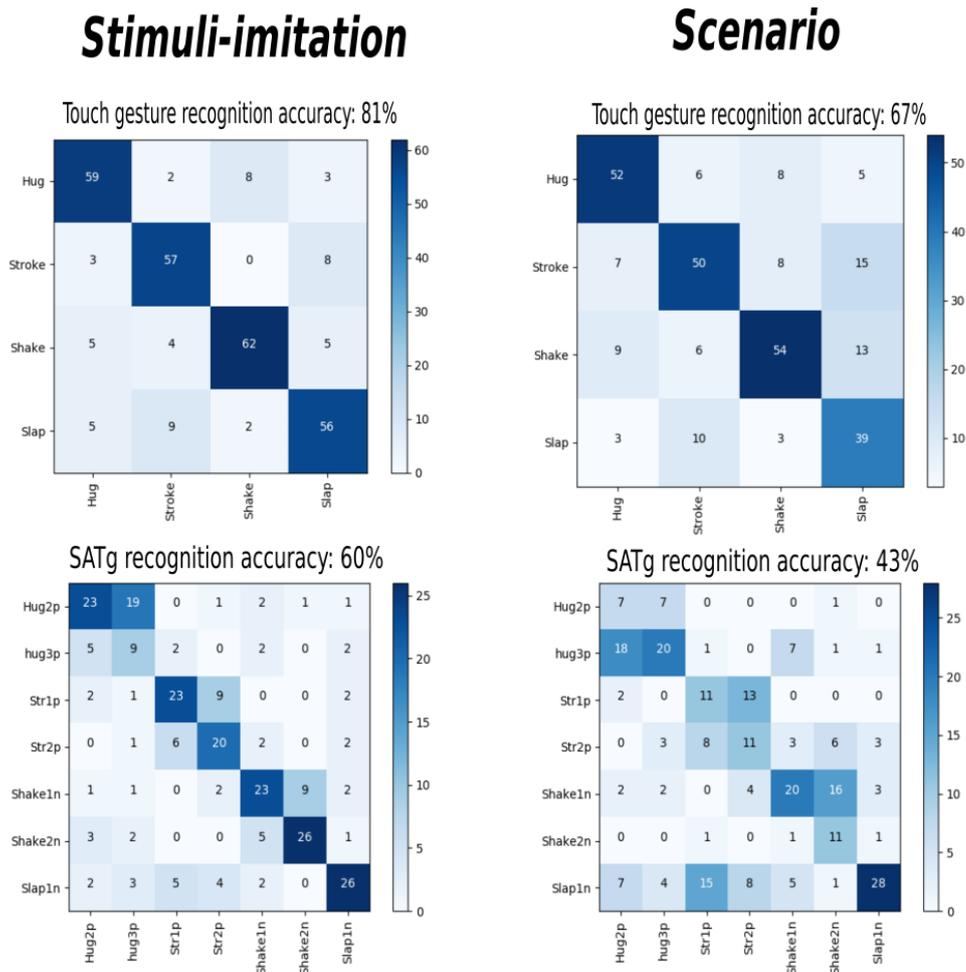


FIGURE 4.13 – Touch gesture and SATg recognition.

4.5 Affect Elicitation through observed Socio-Affective Touch

In this section we measure the influence of contextual information on affect elicitation through observed Socio-affective Touch. The context is encoded in scenarios that evoke a certain touch expression (Table 4.1), and the observed videos are an implementation of the scenarios by actors [89]. Accordingly, two conditions are considered :

- **SI** condition : participant reads the scenario and watch the associated video before assessing their Affect (Context-dependent).
- **SATED** condition : participant watches only the video before assessing their Affect (Context-free).

In the SI condition, 20 participants rated 3 times (repetitions) 14 pairs (video, scenario) in term of valence and arousal. To match SATED annotations where 20 participants rated 2 times the same 14 videos we only considered the 2 first repetitions for SI. Therefore, the total number of annotations per SATg for SI and SATED is 80.

First, to assess the influence of context we compared for both conditions : (i) the annotations consistency ; and (ii) the valence and arousal distributions for each SATg.

Second, we explored also how individual traits such as dispositional empathy and personality modulates affect elicitation in participants.

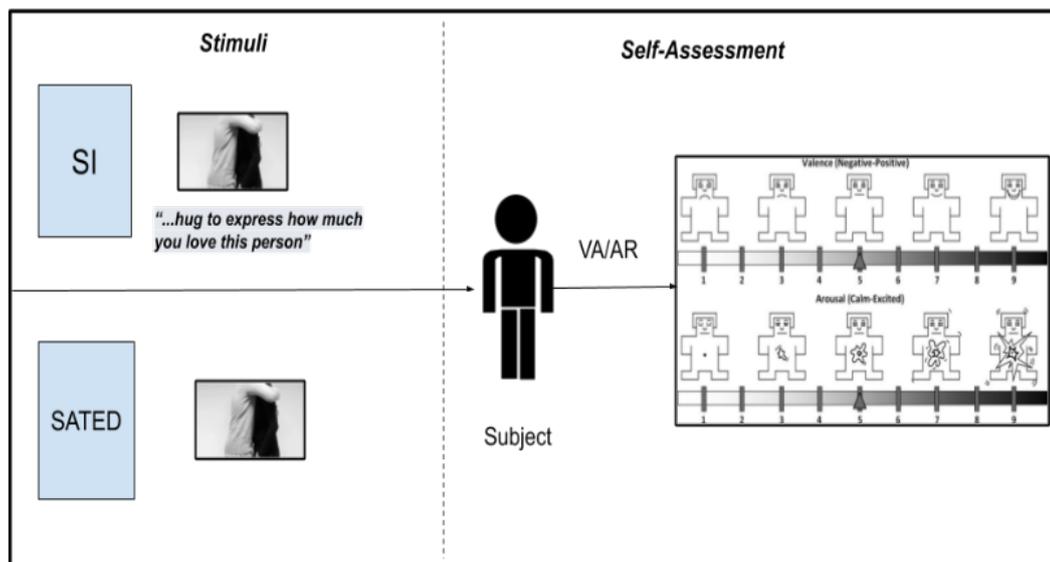


FIGURE 4.14 – Difference between an SI and SATED stimulus in the case of a Hug touch gesture

4.5.1 Annotations consistency

The purpose here is to evaluate the influence of context on annotation consistency by comparing the SI and SATED conditions. We believe that providing participants with a scenario in addition to the video will improve annotation consistency.

Inter-subject consistency : Per participant and for each of the 14 pairs (videos and associated scenarios) we averaged the ratings across the first and second repetition then calculated Spearman rank-order correlation between every pair of participants. The results (Figure 4.15) shows that there is a strong correlation between participants ratings in term of Arousal (Median=0.88, $P < 0.005$) and a moderate correlation for Valence (Median=0.60, $P < 0.05$). Furthermore, a Welch's t-test show a significant mean difference in term of Valence between SI and SATED ($t(378)=3.95$, $P < 0.0001$).

Intra-subject consistency : Per participant we calculated Spearman rank-order correlation of ratings between the first and the second repetition. The results depicted in figure 6 shows that participants were consistent in their ratings of Valence (Median=0.92115, $P < 0.001$) and Arousal (Median=0.78, $P < 0.01$) for the 14 pairs (videos and associated scenario). Furthermore, a Welch's t-test show a significant mean difference in term of Arousal between SI and SATED ($t(38)=4.08$, $P < 0.0005$).

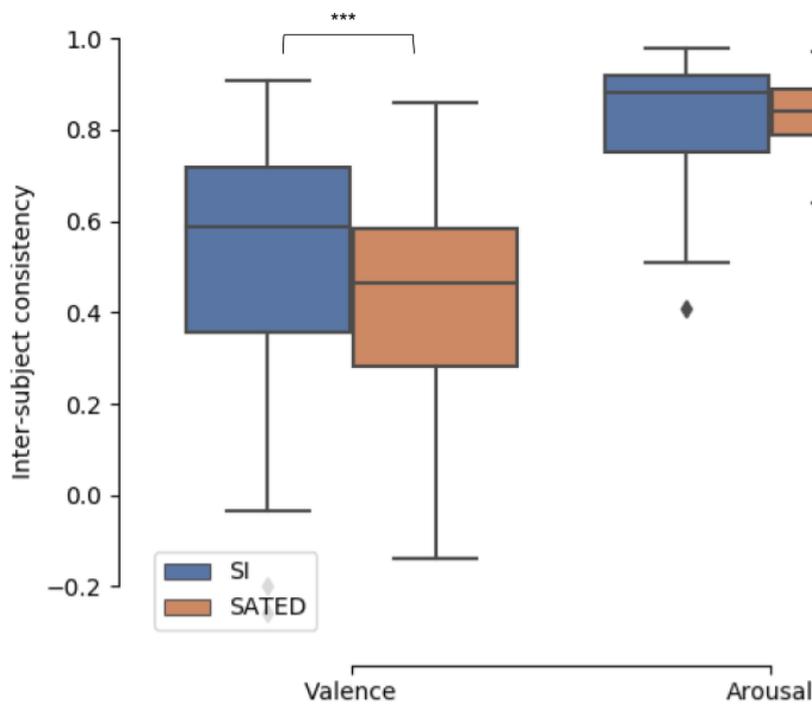


FIGURE 4.15 – Inter-subject consistency

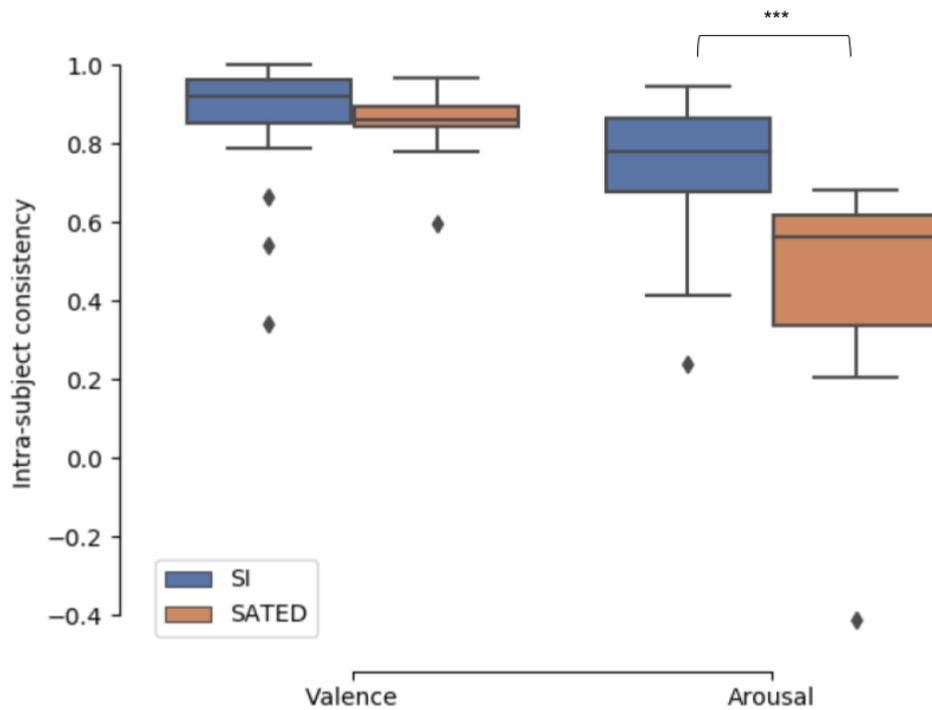


FIGURE 4.16 – Intra-subject consistency

4.5.2 Influence of Context on Valence and Arousal annotations

For valence (Figure 4.17), a Mann–Whitney U test indicates a significant difference between SI and SATED conditions for hug3_p ($U=582$ $P<0.0001$) and str2_p ($U=542$, $P<0.0001$). When the scenario was added, both hug3_p and str2_p valence scores shifted respectively from positive scores (Mean=7.01, SD=1.39) and (Mean=6.75, SD=1.14) to negative scores (Mean=4, SD=1.56) and (Mean=4.19, SD=1.36). Furthermore, SATg that belongs to the same touch gesture such as (hug3_p and hug2_p) and (str1_p and str1_p) were rated as similar for the SATED condition and dissimilar for the SI condition. For Arousal, no significant difference was found between the two conditions.

4.5.3 Interactions with Individual Traits : Empathy and Personality

In this section we used only the SI annotations. For each SATg (e.g hug3_p), we divided the annotation into two groups using the median value of the studied traits (e.g IRI_PT) then applied the Mann-Whitney U test to compare the two groups in terms of Arousal (Table 4.4) and Valence (Table 4.5).

4.5. Affect Elicitation through observed Socio-Affective Touch

For Valence (Table 4.5), significant results were found only for the hug3_p and str2_p SATg. The SATg hug3_p was perceived as more negative by participants that scored higher in perspective taking and empathic concerns. In addition, hug3_p and str2_p were perceived as more positive by participants that scored higher in Openness and Neuroticism. Furthermore, results on Arousal are summarized in Table 4.4. Regarding Extraversion and Conscientiousness no significant results were found.

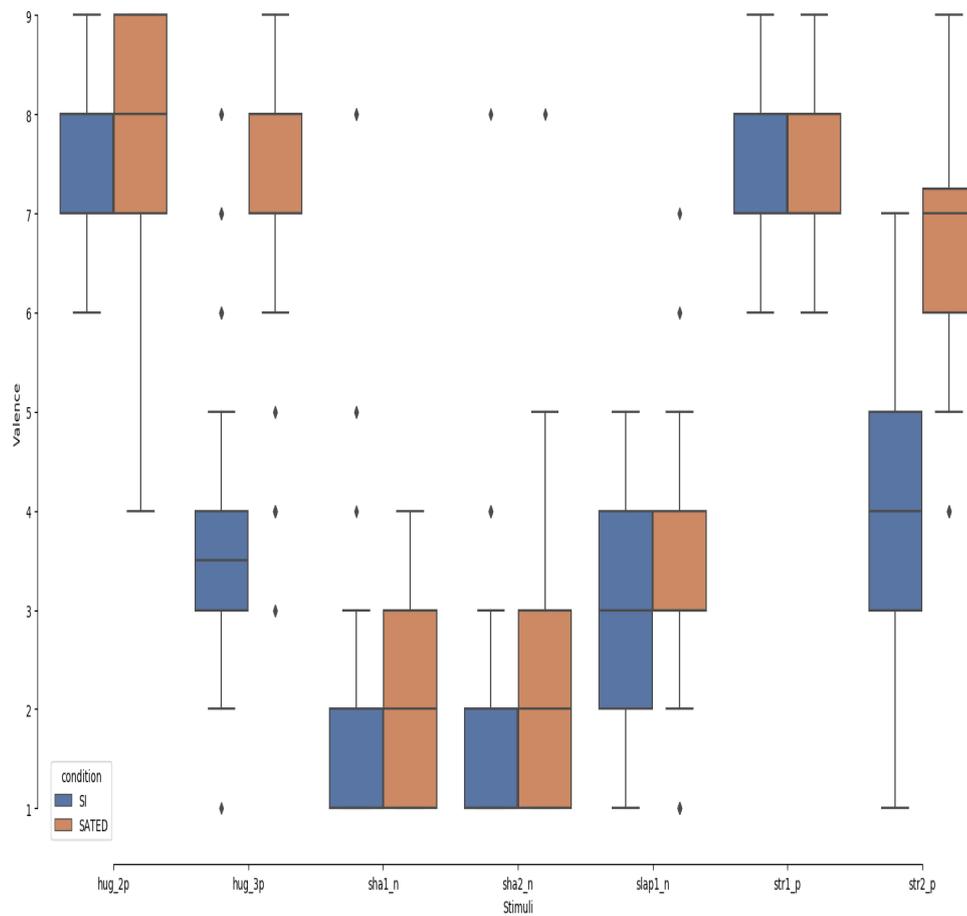


FIGURE 4.17 – Valence annotations for each stimuli in the SI and SATg conditions.

	Arousal				
	SATg	Median ⁺	Median ⁻	U(70)	P
IRI_PT	Hug3_p	4	5	1583	1e-4
	Slap1_n	5	7	1659	3e-4
	Str2_p	3	4	1609	1e-4
IRI_EC	Hug3_p	4	5	1689	7e-4
	Str2_p	3	4	1609	1e-4
IRI_PD	Str1_p	3.5	5	1707	1e-2
Openness	Shake2_n	8	9	1612	2e-3
Neuroticism	Hug2_p	7	5	1904	1e-2

TABLE 4.4 – Influence of empathy and personality on Arousal.

	Valence				
	SATg	Median ⁺	Median ⁻	U(70)	P
IRI_PT	Hug3_p	3	4	1672	4e-4
IRI_EC	Hug3_p	3	4	1800	2e-3
Openness	Hug3_p	4	3	1456	3e-4
	Str2_p	5	3	1342	4e-5
Neuroticism	Hug3_p	4	3	1756	1e-3
	Str2_p	5	4	1755	1e-3

TABLE 4.5 – Influence of empathy and personality on Valence.

4.6 Discussion

Social Touch Recognition : The results on Social Touch Recognition (Figure 4.13) indicates that the model trained on the Stimuli-imitation condition (SI) gave better accuracy scores than the one trained on the S condition ($SVM_{SI}(SI) > SVM_S(S)$). By employing imitation, we constrained SI subjects to follow prototypical exemplars of touch gestures, allowing a reduction of inter-subject variability. It is worth noting that the length of the imitated videos (3 seconds) was sufficient to understand the gist of the interaction, but not long that subjects could replicate exactly the same touch gestures. Therefore, a part of the variability was conserved in order to train models that generalize.

Furthermore, result obtained on SATg (Socio-Affective Touch gesture) recognition (Figure 4.13) shows that models trained on SI were more accurate at differentiating SATg that belong to the same touch gesture (See F1-scores in results section). A straightforward explanation is that subjects in SI were exposed to a specific set of videos for each SATg (e.g hug2p videos are 15 and 28) while subjects in S used their imagination and experience to generate touch gestures.

Also, we could argue that the advantage that SI has on S on both tasks (Touch gesture and SATg recognition) is a consequence of placing the sensors on the mannequin in accordance with the imitated videos (Stimuli selection section). Therefore, SI subjects are more likely to touch more sensors than the S condition.

Although it is intuitive to think that models trained on SI (videos exemplars) will generalize badly on the S (subjective exemplar), results in Tables 4.2 and 4.3 suggest the opposite ($SVM_{SI}(S) \approx SVM_S(S)$). This could mean that the collected SI data are more prototypical and representative of each scenario, or that we need more sensors to capture the complexity of touch gestures in the Scenario condition.

Moreover, we found that a model trained on S gave better (similar) accuracy on SI for Touch gesture (SATg) recognition than on S ($SVM_S(SI) > SVM_S(S)$). This result was unexpected and could mean that SI data are clustered around **prototypes** [111]: *"the most typical example of a concept abstracted from a person experience with a category's exemplars"*. For the S condition prototypes are subjective and depends on the participants experience, while for SI prototypes are the imitated videos from sated (Figure 4.18). Therefore, SI data are easier to classify with a model trained with S data.

Affect Elicitation : The results obtained on Affect elicitation through observed Social Touch (Figure 4.17) suggest that subjects tend to rate as similar SATg that belong to the same touch gesture when context is not available (SATED condition), even though the socio-affective meaning is different. As an example, the hug2_p represent a touch gesture that convey love and intimacy while hug3_p is a consolation hug. From a third person view, a consolation hug is perceived as positive. But when context was introduced in the form of a scenario, in which subjects adopted a first person perspective, for instance, str2_p :*"Your friend is crying. You want to console this person by stroking on his (her) arms."* valence annotations were organized around meaning and affect was experienced vicariously through actors. By using the pronoun 'Your/You' we draw subjects attention towards a specific socio-affective situation; and by displaying an emotionally charged videos we induce a specific emotional state. Consequently, the emotional annotations collected are situation-dependent and reflect an approximation of the subject affective state. Furthermore, findings in Tables 4.5 and 4.4 show that difference between subjects in Personality (Neuroticism and Openness) and Empathy (Perspective Taking, Empathic Concerns and Personal Distress) traits influence how SATg stimuli are perceived.

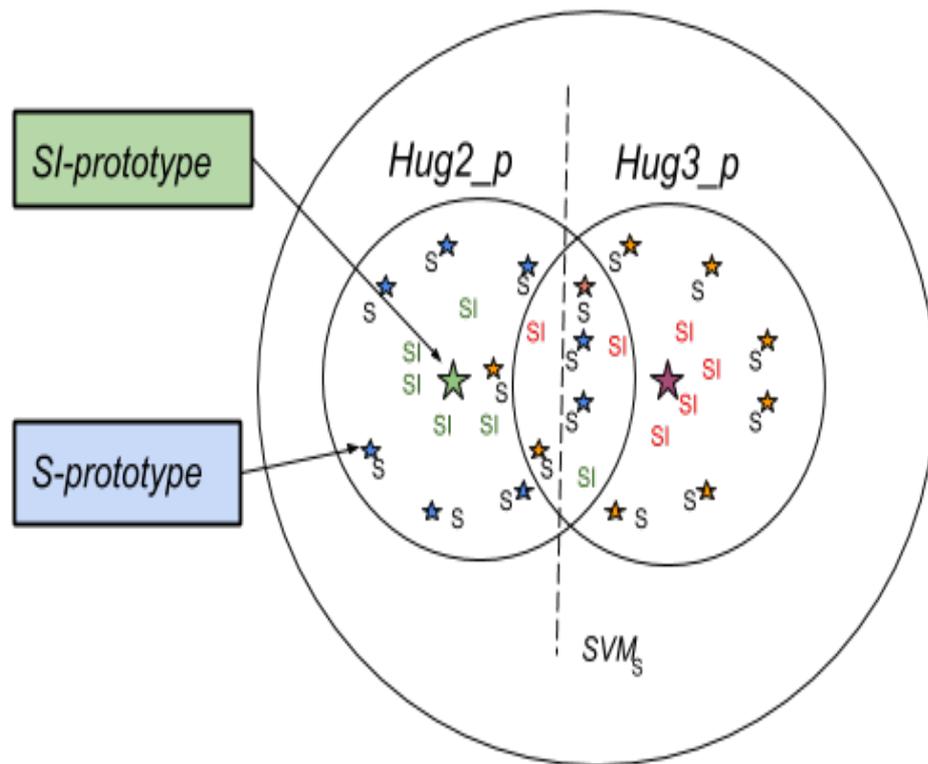


FIGURE 4.18 – A representation of Stimuli-imitation and Scenario prototypes.

Consequently, the collected annotations are idiosyncratic. However, inter-subject (Figure 4.15) and intra-subject (Figure 4.16) consistency results show that subjects were consistent in their annotations. The stimuli-imitation condition subjects were more consistent in their annotations than the ones that belong to the SATED condition. Meaning that providing a context converges subjects annotations towards a shared socio-affective meaning.

4.7 Conclusion

The present study focuses on providing a novel way of collecting Socio-Affective Touch data, a framework that put constraint on Touch gesture elicitation by means of imitation, and modulates Affect elicitation by including context through Socio-Affective Scenarios. Compared to other elicitation methods present in the literature (Table 2.1) where stimuli is vague (e.g tap or love labels), subjective and divorced from social context. Our framework provides subjects with specific socio-affective stimuli (videos and scenarios) then asked them to imitate the stimuli hence the name stimuli-imitation. The framework seeks collecting prototypical data that could be used to train Social Touch Recognition systems that generalize well on various

4.7. Conclusion

conditions - in this study we only tested the Scenario condition. Additionally, the framework reduces the number of required sensors to capture effectively a touch gesture, and consider an interaction that involves the upper-body instead of focusing on a specific body part (e.g. mannequin arm [82]). Furthermore, most Social Touch studies developed systems that successfully recognized Touch gesture [81], Social messages and Emotions [124]. However, neglected modulating factors that influence subject touch expression such as, social context and relationship (Table 2.1). Our framework incorporates these modulators into the data collection process in a manner that reduces subjects cognitive load : instead of remembering or imagining an affective episode, subjects are provided with a scenario and a real life instance (video) to imitate. Our approach is similar to *Stanislavski* active analysis : "The best way to analyse the play, is to take action in the given circumstances", favoring embodiment over intellectualization of the Scenarios (Table 4.1) thus fostering authentic affect expression.

This work represents our effort towards building a comprehensive framework for studying Social Touch in Human-Machine Interaction. A framework that could be used to develop Personalized Affect Recognition system [112] for Social Touch. In the next chapter we will build a personalized model for affect recognition through touch.

Chapter 5

Recognition of Affect and Individual Traits through Touch

5.1 Introduction

Socially intelligent machines are designed to naturally interact with humans [83]. To achieve this goal, machines should be capable of recognizing affect from socio-emotional signals originating from humans. They rely on sensors to gather information from different modality such as vision, audition and touch and uses automatic analysis methods such as Machine Learning to assess affect [116]. Traditionally, these methods map the input features extracted from sensor data with an output that is usually a label (e.g emotion label : "Anger"). However, this approach does not take into consideration individual and cultural differences as opposed to *Personalized* systems [112]. In chapter 4, we found that differences in terms of individual traits influenced valence and arousal annotations, and therefore building a personalized affect recognition system is possible.

The aim of this chapter is to build and evaluate two systems : (i) personalized affect recognition system; and (ii) an individual trait recognition system. These systems will be built upon the stimuli-imitation framework discussed in the previous chapter, and therefore data collected from the SI condition will be used to train machine learning models (Random forest) to make predictions.

The Figure 5.1 describes the personalized affect recognition system. Here we believe that including personal information could enhance the performance of machines in detecting affect from touch. An assumption behind building a personalized system, is that the machine should be acquainted with the human, meaning that personal information is already extracted or that the machine is able to extract these information by itself. In this chapter, personal information was collected via forms (details in *chapter 4*) and concerned two stable traits that change slowly through time : *personality* and dispositional *empathy*. We believe that a personalized system could provide

an attentive care in elderly care centers or hospital, where human contact is the most needed.

In the second system, we use only touch data to predict individual traits such as personality and empathy. We believe that these traits influence how a person performs a touch gesture, and that touch reveals idiosyncratic information about the person that touch.

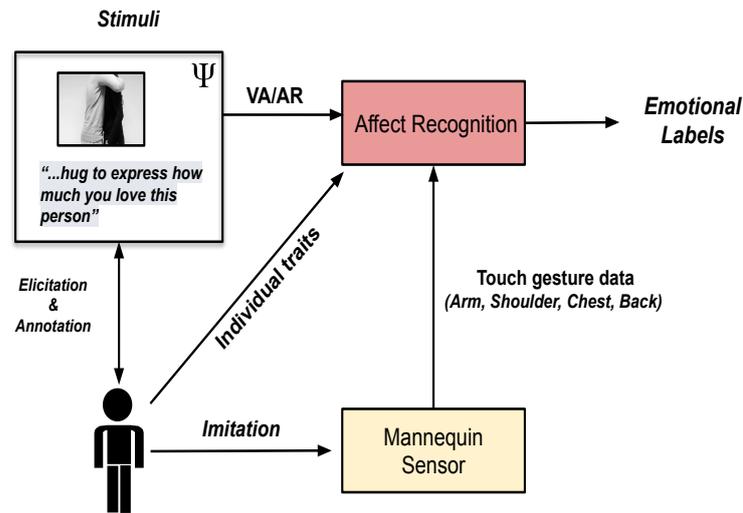


FIGURE 5.1 – Personalized Affect Recognition

5.2 Methodology

5.2.1 Collected data

Socio-affective Touch data

We used socio-affective touch gestures collected through the Stimuli-imitation (SI) method described in *Chapter 4*. The SI stimulus was composed of a written scenario that described the social context and a touch interaction video between actors. First, participants evaluated in term of Valence and Arousal the stimulus using the Self-assessment Manikin [17], then imitated the socio-affective touch gesture, depicted in the video, on a mannequin equipped with 4 sensors that detects pressure and position. Samples were collected from 20 participants with 3 repetitions for 14 videos therefore the total number of touch gestures is 840 (20 participants x 14 videos x 3 repetitions), and each touch gesture was associated to a Valence and Arousal score (Figure 4.9). Furthermore, the sensors were placed respectively on the mannequin Arm, Shoulder, Chest and Back, and collected a 12x12 pressure matrix (144 channels) at a sampling rate of 30 Hz. Features were extracted from these sensors as described in *Chapter 4* (135 features x 4 sensors), consequently we obtained a data set of (840 samples, 540 features).

Individual Traits data

In *Chapter 4* we employed 2 forms to collect Individual Traits data from participants : (i) The Interpersonal Reactivity Index (IRI) measures dispositional empathy, it reflects the ability of a person to adopt others feeling and point of view, and it is a stable character trait. IRI is composed of four distinct subscales Perspective Taking, Empathic concern, Personal distress and Fantasy [34]; (ii) The big Five model proposes a taxonomy of personality along five dimensions : Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism [108]. A French version of each form was given to 20 participants. The Table 5.1 summarizes the collected data.

Individual Traits	Our study	
N	11 Males	9 Females
Fantasy (FS)	15.90 (6.80)	16.55 (6.32)
Perspective Taking (PT)	18.63 (5.37)	17.11 (7.04)
Empathic Concern (EC)	16.90 (5.14)	18.33 (5.26)
Personal Distress (PD)	11.09 (5.06)	9.33 (3.42)
Extraversion (E)	2.95 (0.89)	3.50 (0.84)
Agreeableness (A)	3.88 (0.61)	3.63 (0.97)
Conscientiousness (C)	3.49 (0.71)	3.41 (0.63)
Neuroticism (N)	2.69 (0.74)	3.05 (1.16)
Openness (O)	3.97 (0.52)	3.56 (0.87)

TABLE 5.1 – Mean and standard deviation of Personality dimensions and Empathy subscales.

5.2.2 Labels

Emotional labels

In this study we chose to represent affect using the Russell’s circumplex model [113], a dimensional model in which emotions are located following mainly valence (x-axis) and arousal (y-axis). Similar to Altun and Maclean [4] we used a modification of the of Russell’s model called the *affect grid* [113]. The affect grid (Figure 5.2) divides the space into 9 affective states $\{e_i, i=1..9\}$.

Furthermore, stimuli were annotated by participants in terms of valence and arousal using the self-assessment manikin, where scores ranged from 1 to 9 (Figure 4.9). To represent our stimuli in the affect grid, we associated to each affective state a valence and arousal range as depicted in Figure 5.2. For instance, the scores that fall within (4,5,6) for valence and (7,8,9) for arousal will be mapped to the *aroused* state (e_2). After transforming all valence and arousal scores into emotional labels, we obtain a database that maps touch gesture features with emotional labels, and the Figure 5.3 shows the frequency of each emotional label.

(7, 8, 9)	e_1 distressed	e_2 aroused	e_3 excited
(4, 5, 6)	e_4 miserable	e_5 neutral	e_6 pleased
(1, 2, 3)	e_7 depressed	e_8 sleepy	e_9 relaxed
	(1, 2, 3)	(4, 5, 6)	(7, 8, 9)

FIGURE 5.2 – Location of emotional labels on the Affect-grid [4]

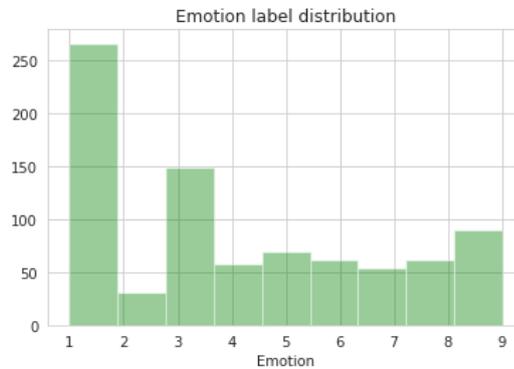


FIGURE 5.3 – Histogram of emotional labels (e_i) obtained from the valence and arousal scores.

Individual Traits labels

Personality and Empathy scores collected in our study (Table 5.1) were divided into two classes (High and Low) for each dimension using mean scores from the original studies [34] [108] (see IRI and BFI Tables in Appendix A). As example, in the original study of IRI [34] the mean value of Perspective Taking (PT) for Males is 16.78. Consequently, Males in our study with a PT score higher than 16.78 will be labeled as high in PT. In this task, classification is binary (High or Low) and the evaluation is done using the average of f1 scores and accuracy.

5.2.3 Loss functions

To evaluate the personalized affect recognition classifier, we introduce two loss functions. The first is the discrete loss function (L_D) defined such that a classification is either correct or incorrect, for instance, mis-classifying e_1 (distressed) with e_9 (relaxed) is similar to mis-classifying e_1 (distressed) with e_2 (aroused).

$$L_D(e_j; e_i) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

The second is the Euclidean loss function (L_E), defined in equation 5.1 such as opposite labels of the affect-grid diagonal (e.g e_1, e_9) yield a loss value of 1 (As a consequence the value of k is $\frac{1}{\sqrt{8}}$). L_E is a distance-based loss function that assigns a degree to an incorrect classification, and therefore take into consideration the proximity of the predicted emotional label e_i to the true emotional label in the affect-grid (*near misses*).

$$L_E(e_j; e_i) = k\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (5.1)$$

Both of these loss functions were introduced in Altun and MacLean paper [4] (*classifier evaluation based on a choice of loss functions section*) to evaluate the classifier performance. Table 5.2 is an example of the L_E loss function applied on the e_1 emotional state. It shows that mis-classifying e_1 with e_2 gives a loss value of 0.35 and mis-classifying e_1 with e_9 gives a loss value of 1. As opposed to the L_D loss function which always yield a loss value of 1, the L_E loss function takes into account the closeness of the affective states in the affect-grid (Figure 5.2).

$L_E(e_1; e_1)=0$	$L_E(e_2; e_1)=0.35$	$L_E(e_3; e_1)=0.71$
$L_E(e_4; e_1)=0.35$	$L_E(e_5; e_1)=0.5$	$L_E(e_6; e_1)=0.79$
$L_E(e_7; e_1)=0.71$	$L_E(e_8; e_1)=0.79$	$L_E(e_9; e_1)=1$

TABLE 5.2 – Euclidean loss for e_1

The expected loss \mathcal{L}_i calculates the total loss for mis-classifying e_i and is defined as follow :

$$\mathcal{L}_i(L) = \sum_{j=1}^9 L(e_j; e_i)P(e_j; e_i); i=1..9$$

Where $P(e_j;e_i)$ is the probability of predicting e_j when the truth is e_i . This probability is calculated from the the classifier predictions (confusion matrix), and is defined as follow :

$$P(e_j;e_i) = \frac{\text{Number of samples } e_i \text{ classified as } e_j}{\text{Number of samples in class } i}$$

5.3 Analysis and discussion

5.3.1 Training

We used Random Forest (RF) classifier for its ability to handle a large number of features. In addition, the scikit-learn implementation of RF has a built-in method to measure feature importance based on gini impurity which is useful for comparing and evaluating features. Also, we saw in *chapter 3* that RF is effective for Touch gesture recognition tasks [133] and outperforms other methods such as support vector machine (SVM). The emotional labels (Figure 5.3), and Individual Traits labels distributions are unbalanced, therefore we used a 10-fold *Stratified* cross validation from scikit-learn [104] where folds were made by preserving the percentage of samples for each class. On the train split we used an up-sampling technique called SMOTE [26] to enhance the performance of the classifier. However, the validation/test split was left untouched to avoid bias. Furthermore, we used the the dummy classifier (scikit-learn) as baseline to compare with the Random Forest. We tested 3 strategies for the dummy classifier : (i) *stratified* : generates predictions by respecting the training set's class distribution; (ii) *most_frequent* : always predicts the most frequent label in the training set; (iii) *uniform* : generates predictions uniformly at random. Lastly, data was standardized (subtract mean and divide by standard deviation).

5.3.2 Personalized Affect Recognition

Influence of Individual Traits on the classifier performance

Random Forest models were trained on two different features. The first model (RF_X) was trained and tested using only touch data, while the other (RF_{X+IT}) used the combination of touch data and individual traits (IT). The results in Table 5.1 show that both models outperform the baselines (stratified, most_frequent and uniform), and that the RF_{X+IT} give better than RF_X for accuracy (0.51% against 0.44%) and for the average of f1 scores (0.35 against 0.26). Also, we noticed that e_1 f1-score was far better (Mean :0.79 ; Std : 0.03) than e_2 (Mean :0.18 ; Std : 0.17) for RF_{X+IT} . This result could be explained by the fact that we have more e_1 than e_2 instances (Figure 5.3) thus the model find it easier to correctly classify e_1 .

5.3. Analysis and discussion

Furthermore, we measured the influence of all features (Sensors features and individual traits scores combined) by using the Random Forest feature selection module [104]. After pre-processing features importance scores (standardization), we considered only values that are above the mean (positive values). The Figure 5.4 shows that features extracted from the Arm sensor are more important than features extracted from the other sensors (back, chest and shoulder). This result could be explained by the fact that the mannequin was touched in the Arm for most touch gestures (Hug, Stroke, Shake, Slap). Additionally, we found that individual traits comes at the second place in terms feature importance which means that personality and empathy features are discriminative thus enhances the classifier performance.

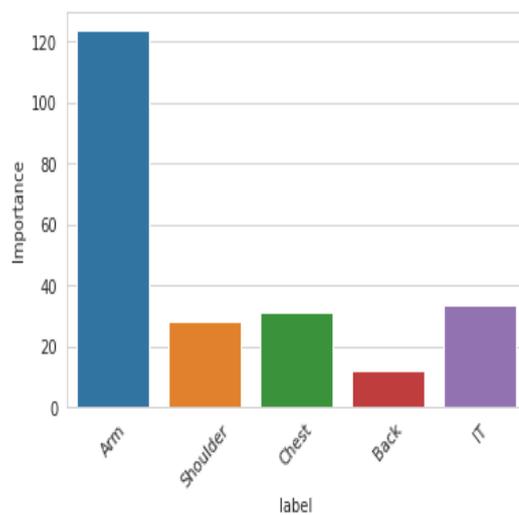


FIGURE 5.4 – Random Forest features importance result.

Classifier	Random Forest		Baselines		
condition	RF_X	RF_{X+IT}	stratified	most_frequent	uniform
e_1	0.77(0.04)	0.79(0.03)	0.29(0.05)	0.48(0.0)	0.17(0.05)
e_2	0.12(0.15)	0.18(0.17)	0.09(0.08)	0.0(0.0)	0.03(0.04)
e_3	0.37(0.03)	0.48(0.05)	0.19(0.05)	0.0(0.0)	0.13(0.06)
e_4	0.22(0.09)	0.33(0.16)	0.12(0.03)	0.0(0.0)	0.13(0.1)
e_5	0.14(0.07)	0.17(0.07)	0.07(0.01)	0.0(0.0)	0.09(0.06)
e_6	0.17(0.14)	0.29(0.1)	0.05(0.04)	0.0(0.0)	0.1(0.03)
e_7	0.2(0.12)	0.21(0.09)	0.05(0.07)	0.0(0.0)	0.06(0.05)
e_8	0.05(0.06)	0.28(0.15)	0.05(0.04)	0.0(0.0)	0.06(0.03)
e_9	0.35(0.11)	0.46(0.1)	0.13(0.09)	0.0(0.0)	0.07(0.02)
Accuracy	0.44(0.03)	0.51(0.03)	0.17(0.03)	0.32(0.0)	0.11(0.02)
f1-scores avg	0.26(0.03)	0.35(0.02)	0.12(0.02)	0.05(0.0)	0.09(0.02)

TABLE 5.3 – Classification results of emotional labels (e_i) reported using Mean and Std of f1-scores obtained from the 10-fold cross validation for different conditions. At the bottom of the table results were summarized using Accuracy and f1scores average.

Evaluation of RF_{X+IT} using loss functions

In the previous section we found that including individual traits enhanced emotional labels classification results (Accuracy : 0.51, f1-avg : 0.35) (Table 5.3). In this section we want to evaluate the RF_{X+IT} classifier in a fine-grained manner, instead of using summary scores (e.g accuracy) we want to inspect mis-classification errors for each emotional label.

In a Human-Machine interaction, the quality of an interaction depends on the ability of the machine to accurately recognize the human affect hence classification errors impact negatively the interaction. However, not all errors are equal, for instance, mis-classifying a relaxed state (e_9) with distressed state (e_1) does not have the same impact on an interaction as mis-classifying a relaxed state (e_9) with a pleased state (e_6). Consequently, in this section, we want to determine, for each emotional label, the closeness of mis-classification.

As we discussed earlier, the L_E loss function penalize highly mis-classification where the predicted label is far from the true label in the affect-grid, while the L_D is indifferent to the closeness of the mis-classification. A first step is to calculate the expected loss for each emotional label. The confusion matrix for RF_{X+IT} (Figure 5.5) was

5.3. Analysis and discussion

calculated by averaging the classification results obtained from the 10-fold cross-validation. We used this matrix to estimate the probability $P(e_j;e_i)$ (probability of predicting e_j when the truth is e_i), then calculated the expected loss $\mathcal{L}_i(L)$ for the Euclidean distance loss L_E and the Discrete loss L_D on each emotional label e_i . Results in Table 5.4 show that $\mathcal{L}_i(L_E) < \mathcal{L}_i(L_D)$ which is an expected result because L_E penalize less close mis-classifications (near misses) than L_D .

However, expected losses $\mathcal{L}_i(L)$ do not take into account the frequency of the emotional labels (Figure 5.3), therefore, we calculated for each emotion the percentage decrease in loss for L_E ($p(E)$) and L_D ($p(D)$) for the RF_{X+IT} and the Stratified classifier. Both $p(E)$ and $p(D)$ were calculated using the equation 5.2, and were used to calculate the *Near-miss rate* as defined in the equation 5.3. A high NM rate for an emotional label e_i means that e_i was mis-classified with proximal emotional labels in the affect-grid (e.g e_1 and e_2).

$$p_i(E) = \frac{\mathcal{L}_i(\text{strat}_E) - \mathcal{L}_i(L_E)}{\mathcal{L}_i(\text{strat}_E)} \quad (5.2)$$

$$NM_i = p_i(E) - p_i(D) \quad (5.3)$$

e_i	$\mathcal{L}_i(L_E)$	$\mathcal{L}_i(\text{strat}_E)$	$p_i(E)$	$\mathcal{L}_i(L_D)$	$\mathcal{L}_i(\text{strat}_D)$	$p_i(D)$
1	0.16	0.45	0.64	0.30	0.70	0.57
2	0.29	0.48	0.39	0.53	1.00	0.47
3	0.34	0.56	0.39	0.54	0.84	0.35
4	0.33	0.39	0.15	0.70	0.86	0.18
5	0.35	0.41	0.14	0.82	0.89	0.07
6	0.29	0.57	0.49	0.71	0.92	0.22
7	0.42	0.72	0.41	0.73	0.96	0.24
8	0.32	0.46	0.30	0.64	0.93	0.31
9	0.34	0.66	0.48	0.56	0.89	0.37
Average	0.32	0.52	0.37	0.61	0.89	0.30

TABLE 5.4 – Percentage decrease in loss for the Euclidean and Discrete loss functions calculated for each emotion

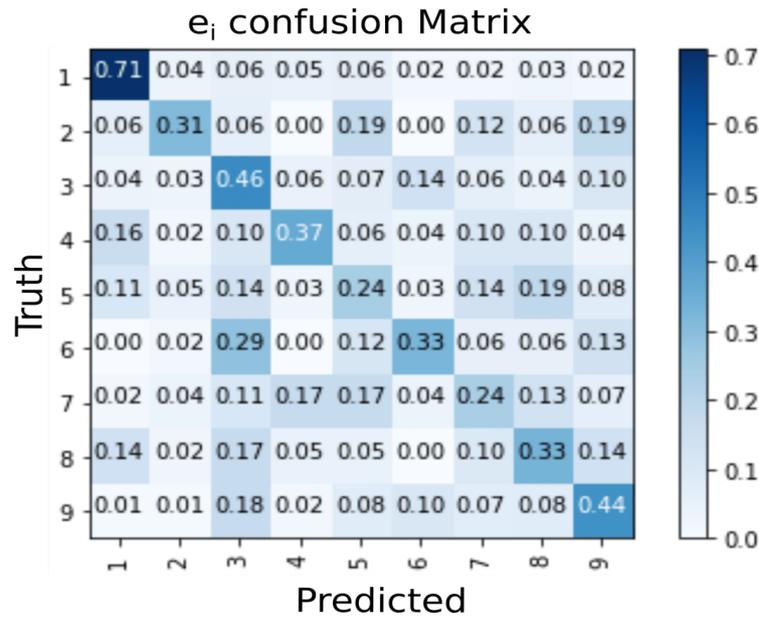


FIGURE 5.5 – Emotional labels confusion matrix.

Discrete miss-classification :

The Figure 5.6 is a bar plot of $p_i(D)$ scores (after subtracting the mean) taken from Table 5.4. It shows that the classifier performed poorly on e_4 , e_5 and e_6 (medium arousal) compared to the e_1 , e_2 and e_3 (high arousal) emotional labels. As for low arousal labels, correct classification were in between high and medium arousal labels. Though Figure 5.6 gives us a fine-grained view of mis-classification, it does not inform us regarding the degree of mis-classification. In the next section we will look into the nature of mis-classification using *near-misses* rates.

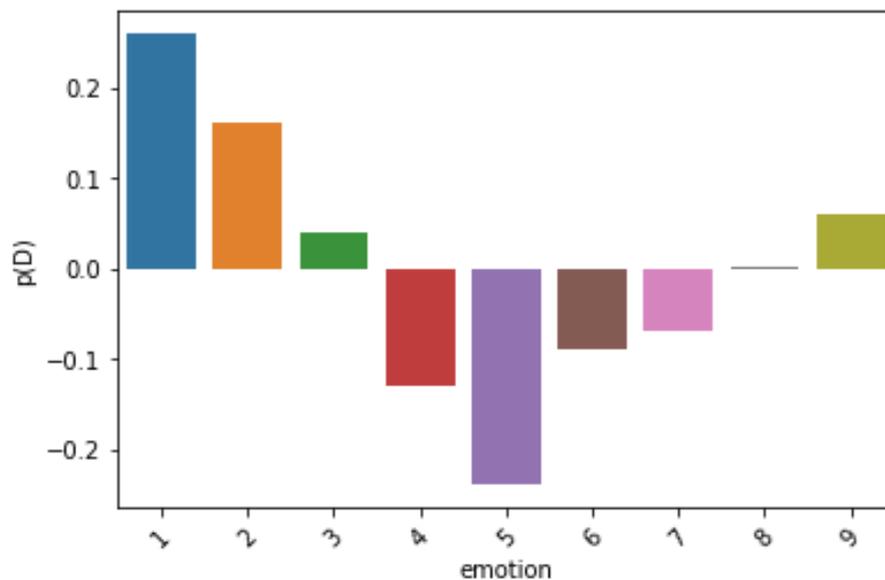


FIGURE 5.6 – Percentage decrease in loss $p(D)$ on each emotion.

5.3. Analysis and discussion

Near misses :

Figure 5.7 depicts the NM rate for each emotion label. We notice that near misses are the highest for e_6 (mostly hug2_p and str1_p) and e_9 (mostly hug2_p and str1_p) in the case of positive valence emotions ; and for e_7 (mostly hug3_p and str2_p) for negative valence emotions.

Looking into the confusion matrix, we notice that e_6 (pleased) and e_9 (relaxed) were mainly confused with e_3 (excited). All the aforementioned emotions shared hug2_p and str1_p as predominant stimuli (SATg Table 4.1) which could explain the confusion : it is difficult for the RF_{X+IT} to separate, for instance, two hug touch gesture that were annotated differently. In addition, we know that subjects imitated the same videos (from SATED) thus it is more likely that confusion occurred as a consequence of annotation variability between subjects, a variability that could be explained by differences in term of individual traits. As we saw in *chapter 4 : Table 4.4*, Perspective taking and Neuroticism scores influenced highly arousal annotations for str1_p and hug2_p. Here, confusion occurred mainly following the arousal direction for e_6 , e_3 and e_9 .

Similarly, we noticed that e_7 (depressed) was confused mainly with e_4 , e_5 and e_8 , and that the predominant stimuli were hug3_p and str2_p. In addition to arousal, valence annotations were also affected by individual traits (Table 4.4 and 4.5), which means that confusion occurred in both direction (valence and arousal).

Although we included Individual Traits as input for our system, it is difficult to eradicate completely confusion, especially for near misses. However, results found in Table 5.1 shows how including IT data mitigated partly this problem.

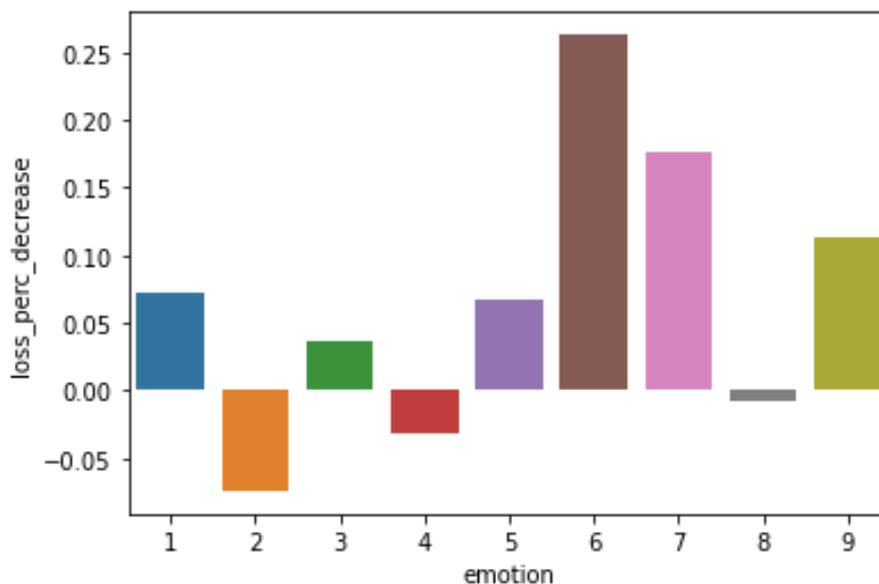


FIGURE 5.7 – Near misses rate for each emotion.

5.3.3 Individual Traits recognition from Touch

In the previous section we found out that affect could be derived from touch in a Human-Machine interaction. Similarly, we want to explore whether information about personality and empathy could be extracted from touch. From the 20 participants that we recruited, individual traits were obtained using questionnaires and concerned only traits that are stable over time. We attributed to each participant a label for each traits (High or Low) using mean scores obtained from studies with large samples [34] [108]. In other words, a participant with a value higher than the mean for a trait (e.g perspective taking) will be labeled as *High* for this trait. The task is a binary classification that follows the same training process as we saw the previous section.

Results for both empathy and personality are summarized in Table 5.5 and 5.6. To evaluate classification performance, we compared our Random Forest classifier with the following baseline : (i) stratified, (ii) frequent and (iii) uniform using accuracy and f1-average score (average of High and Low) and only best baselines results were reported. For both traits we obtained above chance level results.

Empathy	Classifier	High	Low	accuracy	f1-avg
Fantasy (FS)	RF_X	0.77(0.02)	0.59(0.07)	0.71(0.03)	0.68(0.04)
	stratified	0.58(0.03)	0.4(0.06)	0.51(0.04)	0.49(0.05)
	frequent	0.75(0.0)	0.0(0.0)	0.6(0.0)	0.38(0.0)
Perspective	RF_X	0.64(0.04)	0.65(0.03)	0.65(0.03)	0.65(0.03)
Taking (PT)	uniform	0.52(0.03)	0.51(0.03)	0.52(0.02)	0.52(0.01)
Empathic	RF_X	0.71(0.04)	0.6(0.04)	0.66(0.03)	0.66(0.03)
Concern (EC)	frequent	0.71(0.0)	0.0(0.0)	0.55(0.0)	0.35(0.0)
	uniform	0.49(0.03)	0.46(0.02)	0.48(0.01)	0.48(0.01)
Personal	RF_X	0.4(0.07)	0.9(0.01)	0.83(0.02)	0.65(0.03)
Distress (PD)	stratified	0.22(0.05)	0.79(0.01)	0.67(0.01)	0.5(0.03)
	frequent	0.0(0.0)	0.89(0.0)	0.8(0.0)	0.44(0.0)

TABLE 5.5 – Empathy classification results reported with f1-scores.

5.3. Analysis and discussion

BFI	Classifier	High	Low	accuracy	f1-avg
Extraversion	RF_X	0.65(0.06)	0.68(0.05)	0.67(0.05)	0.66(0.06)
	uniform	0.52(0.03)	0.5(0.02)	0.51(0.02)	0.51(0.02)
Agreeableness	RF_X	0.65(0.04)	0.73(0.03)	0.7(0.03)	0.69(0.03)
	stratified frequent	0.43(0.04)	0.54(0.04)	0.49(0.03)	0.48(0.03)
		0.0(0.0)	0.71(0.0)	0.55(0.0)	0.35(0.0)
Conscientiousness	RF_X	0.52(0.04)	0.74(0.04)	0.66(0.04)	0.63(0.04)
	stratified frequent	0.41(0.05)	0.59(0.03)	0.52(0.04)	0.5(0.04)
		0.0(0.0)	0.75(0.0)	0.6(0.0)	0.38(0.0)
Neuroticism	RF_X	0.72(0.02)	0.5(0.05)	0.64(0.03)	0.61(0.04)
	stratified frequent	0.61(0.03)	0.41(0.03)	0.53(0.03)	0.51(0.02)
		0.75(0.0)	0.0(0.0)	0.6(0.0)	0.38(0.0)
Openness	RF_X	0.43(0.04)	0.75(0.02)	0.66(0.02)	0.59(0.03)
	stratified frequent	0.4(0.03)	0.67(0.04)	0.57(0.04)	0.53(0.03)
		0.0(0.0)	0.79(0.0)	0.65(0.0)	0.39(0.0)

TABLE 5.6 – Personality classification results reported with f1-scores.

5.4 Conclusion

In this chapter we showed that adding personal information enhanced affect recognition scores. The affective labels we used in this study were derived from valence and arousal scores annotated by participants. And as we saw in Tables 4.5 and 4.4 (*chapter 4*), participants valence and arousal annotations were influenced by personality traits, therefore including these personal traits as input allowed the model to take into account the differences between participants in terms of stimuli annotation. The results found in the *Near misses* section show that even though including IT improves the model performance (Table 5.1), close labels are difficult to classify. As for individual traits recognition, we obtained above chance level classification of both BFI and Empathy. We believe that performance could be enhanced if emotional labels were balanced thus in future work we will design stimuli such that every emotional label is represented equally.

Chapter 6

Conclusion and Perspectives

This dissertation aims at developing socially intelligent machines capable of interacting with humans through touch. In a Human-Machine interaction, machines are expected to touch and to respond adequately to it. A first step towards building systems capable of achieving this goal, is to mimic how humans interpret and respond to social touch. As we reported in *chapter 2*, the interpretation of interpersonal touch for humans is a result of physiological (bottom-up) and psychological processes (top-down) working in tandem. While in the first, researchers were interested in how information flows from the skin to the brain, in the second there was an interest in studying contextual factors that modulates the meaning of a touch. For socially intelligent machines, the interactive cycle is composed of four successive components : (i) sensing ; (ii) automatic recognition and interpretation ; and (iii) generating or responding to touch. In this thesis we were mainly focused on facing challenges relating the first and second components.

In the case of automatic recognition of touch, we worked in *chapter 3*, towards developing an End-End touch gesture recognition algorithm, that we called *Attentive Touch Network (ATM)*. ATM was used to classify 14 touch gestures for the CoST database (Corpus of Social Touch) and 7 touch gestures for the HAART database (Human-Animal Affective Robot Touch) following the rules dictated by the ICMI Social Touch Challenge. The results we obtained in this challenge (60.9% for CoST ; 67.8% for HAART) were similar to the ones obtained by the winner of the challenge (61.3% for CoST ; 70.9% for HAART). Although we did not outperform the best results, our method required no feature engineering efforts ; and used self-attention as an efficient alternative to Recurrent neural network (RNN) for touch gesture sequence processing.

The social touch challenge offered an opportunity to understand how touch gestures were elicited in a Human-Machine interaction. Using labels to elicit touch gesture presented several limitations : data collected through labels-based approach is context-free, socially divorced and exhibit high inter-subject variability. In *chapter 4* we proposed a novel framework for data collection that take into consideration the synergy between contextual factors (e.g relationship) and touch. We called our framework stimuli-imitation, and it used a scenario-based stimuli enriched with social touch interaction videos. In Stimuli-imitation, a participant reads the scenario and imitates the videos (taken from SATED database). The framework was evaluated on its ability to produce consistent annotations (valence and arousal annotations), and consistent touch gestures. Furthermore, the majority of previous studies considered a localized form of interactions where participants touch, for instance, a mannequin-arm or a zoomorphic animal. In our framework, we involved the full upper-body of participants by selecting touch gestures such as Hug from the SATED database, and by building a sensor (mannequin-sensor) that collect touch data from four body parts (Arm, Shoulder, Chest and Back). In this chapter, we also showed that individual traits such personality and empathy influenced participants valence and arousal annotations of stimuli, and therefore building a personalized affect recognition system is possible.

As we saw in *chapter 2*, an affect recognition system could be built in two ways : (i) one-step system; and (ii) two-step system. The two-step system identifies first the touch gesture then attach an affective meaning to it, while the one-step system derive affective meaning directly from touch sensors data. In *chapter 5* we built a one-step system for personalized affect recognition. The system combined individual traits, touch gesture and annotations (valence and arousal) data, collected under the stimuli-imitation framework, to predict affect. We used a random forest model and a similar set of features as *chapter 4*. The results found, showed that including individual traits enhanced classification performance for accuracy and f1 average scores (summary scores). However, not all classification errors are equal in terms of their impact on the unfolding of an interaction. Consequently, a fine grained analysis that take into consideration the closeness of the affective states in the affect-grid (*near misses*) was deemed necessary for a comprehensive evaluation of classifier (Random forest). Furthermore, we also showed that individual traits could be recognized from touch in an above chance level.

The purpose of this thesis is the integration of socio-affective touch into interactive systems. In the context of Human-Machine interaction, a socially intelligent machine should be able to sense, recognize and respond appropriately to Social Touch. In this research we introduced a novel framework for collecting socio-affective touch data and developed algorithms that enables its interpretation. Therefore, our contributions were mainly about the interpretation of touch from the machine perspective. However, we believe that closing the interactive loop by designing a system that generate adequate touch expressions is the next natural step toward deploying socially intelligent machines in naturalistic environments such as hospitals and healthcare centers. For such applications, it is important to consider challenges such as, designing an evaluation metric that measures the impact of an interaction on a patient (e.g engagement); real-time processing of touch data; and multi-modality by including other channels of communication (e.g voice). Another challenge that is specific to modeling touch is the scarcity of data. As a consequence, it is difficult for machine learning algorithm to find patterns and generalize well. A simple solution is collect more data. However, data collection is costly and time consuming. We believe that in future research we will be working on an overarching framework that could homogenize the data collection process and encompass findings from various disciplines.

Bibliography

- [1] Martín ABADI et al. « TensorFlow : A System for Large-Scale Machine Learning ». In : *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI'16. Savannah, GA, USA : USENIX Association, 2016, 265–283. ISBN : 9781931971331.
- [2] Rochelle ACKERLEY et al. « Touch perceptions across skin sites : differences between sensitivity, direction discrimination and pleasantness ». In : *Frontiers in Behavioral Neuroscience* (2014). DOI : [10.3389/fnbeh.2014.00054](https://doi.org/10.3389/fnbeh.2014.00054).
- [3] Saad ALBAWI et al. *Social Touch Gesture Recognition Using Convolutional Neural Network*. en. Research Article. Oct. 2018. DOI : <https://doi.org/10.1155/2018/6973103>.
- [4] Kerem ALTUN et Karon E. MACLEAN. « Recognizing affect in human touch of a robot ». In : *Pattern Recognition Letters*. Pattern Recognition in Human Computer Interaction 66 (nov. 2015), p. 31-40. ISSN : 0167-8655. DOI : [10.1016/j.patrec.2014.10.016](https://doi.org/10.1016/j.patrec.2014.10.016).
- [5] Mehdi AMMI et al. « Haptic human-robot affective interaction in a handshaking social protocol ». In : *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 2015, p. 263-270.
- [6] Brenna D. ARGALL et Aude G. BILLARD. « A survey of Tactile Human–Robot Interactions ». In : *Robotics and Autonomous Systems* 58.10 (oct. 2010), p. 1159-1176. ISSN : 0921-8890. DOI : [10.1016/j.robot.2010.07.002](https://doi.org/10.1016/j.robot.2010.07.002).
- [7] Malika AUVRAY, Erik MYIN et Charles SPENCE. « The sensory-discriminative and affective-motivational aspects of pain ». In : *Neuroscience & Biobehavioral Reviews* 34.2 (2010), p. 214-223.
- [8] Dzmitry BAHDANAU, Kyunghyun CHO et Yoshua BENGIO. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2014. arXiv : [1409.0473 \[cs.CL\]](https://arxiv.org/abs/1409.0473).
- [9] Sunhye BAI, Rena L REPETTI et Jacqueline B SPERLING. « Children’s expressions of positive emotion are sustained by smiling, touching,

- and playing with parents and siblings : A naturalistic observational study of family life. » In : *Developmental psychology* 52.1 (2016), p. 88.
- [10] Jeremy N. BAILENSEN et al. « Virtual Interpersonal Touch : Expressing and Recognizing Emotions Through Haptic Devices ». In : *Hum.-Comput. Interact.* 22.3 (août 2007), p. 325-353. ISSN : 0737-0024.
- [11] Tugce BALLI ALTUGLU et Kerem ALTUN. « Recognizing Touch Gestures for Social Human-Robot Interaction ». en. In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*. Seattle, Washington, USA : ACM Press, 2015, p. 407-413. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2830600](https://doi.org/10.1145/2818346.2830600).
- [12] K. BARNETT. « A theoretical construct of the concepts of touch as they relate to nursing ». eng. In : *Nursing Research* 21.2 (avr. 1972), p. 102-110. ISSN : 0029-6562.
- [13] Dean C BARNLUND. *Public and private self in Japan and the United States : Communicative styles of two cultures*. Simul Press Tokyo, 1975.
- [14] Alan M BECK et Aaron Honori KATCHER. *Between pets and people : The importance of animal companionship*. Purdue University Press, 1996.
- [15] H. BEKKERING, A. WOHLSCHLÄGER et M. GATTIS. « Imitation of gestures in children is goal-directed ». eng. In : *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 53.1 (fév. 2000), p. 153-164. ISSN : 0272-4987. DOI : [10.1080/713755872](https://doi.org/10.1080/713755872).
- [16] James BERGSTRÄ et Yoshua BENGIO. « Random Search for Hyper-Parameter Optimization ». In : *J. Mach. Learn. Res.* 13.null (fév. 2012), 281–305. ISSN : 1532-4435.
- [17] Margaret M BRADLEY et Peter J LANG. « Measuring emotion : the self-assessment manikin and the semantic differential ». In : *Journal of behavior therapy and experimental psychiatry* 25.1 (1994), p. 49-59.
- [18] Margaret M. BRADLEY et Peter J. LANG. « The International Affective Picture System (IAPS) in the study of emotion and attention ». In : *Handbook of emotion elicitation and assessment*. Series in affective science. New York, NY, US : Oxford University Press, 2007, p. 29-46. ISBN : 978-0-19-516915-7.
- [19] Judee K. BURGOON. « Relational message interpretations of touch, conversational distance, and posture ». In : *Journal of Nonverbal Behavior* 15.4 (1991), 233–259. ISSN : 0191-5886. DOI : [10.1007/bf00986924](https://doi.org/10.1007/bf00986924).
- [20] Judee K. BURGOON et al. « Relational Messages Associated with Nonverbal Behaviors ». en. In : *Human Communication Research* 10.3 (mar.

- 1984), p. 351-378. ISSN : 1468-2958. DOI : [10.1111/j.1468-2958.1984.tb00023.x](https://doi.org/10.1111/j.1468-2958.1984.tb00023.x).
- [21] John-John CABIBIHAN, Lihao ZHENG et Chin Kiang Terence CHER. « Affective tele-touch ». In : *International Conference on Social Robotics*. Springer. 2012, p. 348-356.
- [22] R. A. CALVO et S. D'MELLO. « Affect Detection : An Interdisciplinary Review of Models, Methods, and Their Applications ». In : *IEEE Transactions on Affective Computing* 1.1 (jan. 2010), p. 18-37. ISSN : 1949-3045. DOI : [10.1109/T-AFFC.2010.1](https://doi.org/10.1109/T-AFFC.2010.1).
- [23] Xi Laura CANG et al. « Different Strokes and Different Folks : Economical Dynamic Surface Sensing and Affect-Related Touch Recognition ». In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ICMI '15. Seattle, Washington, USA : Association for Computing Machinery, nov. 2015, p. 147-154. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2820756](https://doi.org/10.1145/2818346.2820756).
- [24] Carissa J. CASCIO, David MOORE et Francis MCGLONE. « Social touch and human development ». In : *Developmental Cognitive Neuroscience* (avr. 2018). ISSN : 1878-9293. DOI : [10.1016/j.dcn.2018.04.009](https://doi.org/10.1016/j.dcn.2018.04.009).
- [25] Jonathan CHANG, Karon MACLEAN et Steve YOHANAN. « Gesture Recognition in the Haptic Creature ». en. In : *Haptics : Generating and Perceiving Tangible Sensations*. Sous la dir. d'Astrid M. L. KAPPERS et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, p. 385-391. ISBN : 978-3-642-14064-8.
- [26] Nitesh V. CHAWLA et al. « SMOTE : Synthetic Minority over-Sampling Technique ». In : *J. Artif. Int. Res.* 16.1 (juin 2002), 321-357. ISSN : 1076-9757.
- [27] H. T. CHUGANI et al. « Local brain functional activity following early deprivation : a study of postinstitutionalized Romanian orphans ». eng. In : *NeuroImage* 14.6 (déc. 2001), p. 1290-1301. ISSN : 1053-8119. DOI : [10.1006/nimg.2001.0917](https://doi.org/10.1006/nimg.2001.0917).
- [28] M. D. COONEY, S. NISHIO et H. ISHIGURO. « Recognizing affection for a touch-based interaction with a humanoid robot ». In : *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2012, p. 1420-1427. DOI : [10.1109/IRoS.2012.6385956](https://doi.org/10.1109/IRoS.2012.6385956).
- [29] M. D. COONEY et al. « Full-body gesture recognition using inertial sensors for playful interaction with small humanoid robot ». In : *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2010, p. 2276-2282. DOI : [10.1109/IRoS.2010.5650081](https://doi.org/10.1109/IRoS.2010.5650081).

- [30] Martin D. COONEY, Shuichi NISHIO et Hiroshi ISHIGURO. « Importance of Touch for Conveying Affection in a Multimodal Interaction with a Small Humanoid Robot ». In : *I. J. Humanoid Robotics* 12 (2015). DOI : [10.1142/S0219843615500024](https://doi.org/10.1142/S0219843615500024).
- [31] Daniela CORBETTA et Winona SNAPP-CHILDS. « Seeing and touching : The role of sensory-motor experience on the development of infant reaching ». In : *Infant Behavior and Development* 32.1 (2009), 44–58. ISSN : 0163-6383. DOI : [10.1016/j.infbeh.2008.10.004](https://doi.org/10.1016/j.infbeh.2008.10.004).
- [32] Ilona CROY et al. « Interpersonal stroking touch is targeted to C tactile afferent activation ». In : *Behavioural Brain Research* *Series A* 297 (2016), p. 37-40.
- [33] April H. CRUSCO et Christopher G. WETZEL. « The Midas Touch : The Effects of Interpersonal Touch on Restaurant Tipping ». en. In : *Personality and Social Psychology Bulletin* 10.4 (déc. 1984), p. 512-517. ISSN : 0146-1672. DOI : [10.1177/0146167284104003](https://doi.org/10.1177/0146167284104003).
- [34] Mark H DAVIS. « Measuring individual differences in empathy : Evidence for a multidimensional approach. » In : *Journal of personality and social psychology* 44.1 (1983), p. 113.
- [35] Quentin DEBARD et al. « Learning to recognize touch gestures : recurrent vs. convolutional features and dynamic sampling ». en. In : IEEE, mai 2018, p. 114-121. DOI : [10.1109/FG.2018.00026](https://doi.org/10.1109/FG.2018.00026).
- [36] Sam DORROS, Alesia HANZAL et Chris SEGRIN. « The Big Five personality traits and perceptions of touch to intimate and nonintimate body regions ». In : *Journal of Research in personality* 42.4 (2008), p. 1067-1073.
- [37] R.I.M. DUNBAR. « The social role of touch in humans and primates : Behavioural function and neurobiological mechanisms ». In : *Neuroscience Biobehavioral Reviews* 34.2 (2010), 260–268. ISSN : 0149-7634. DOI : [10.1016/j.neubiorev.2008.07.001](https://doi.org/10.1016/j.neubiorev.2008.07.001).
- [38] Eric EICH et al. « Combining music with thought to change mood ». In : *Handbook of emotion elicitation and assessment*. Series in affective science. New York, NY, US : Oxford University Press, 2007, p. 124-136. ISBN : 978-0-19-516915-7.
- [39] Mohamad A. EID et Hussein AL OSMAN. « Affective Haptics : Current Research and Future Directions ». In : *IEEE Access* 4 (2016), p. 26-40. ISSN : 2169-3536. DOI : [10.1109/ACCESS.2015.2497316](https://doi.org/10.1109/ACCESS.2015.2497316).
- [40] Paul EKMAN et Wallace V. FRIESEN. *Unmasking the face : A guide to recognizing emotions from facial clues*. Unmasking the face : A guide to

- recognizing emotions from facial clues. Oxford, England : Prentice-Hall, 1975. ISBN : 978-0-13-938183-6.
- [41] Wail EL BANI et Mohamed CHETOUANI. « Touch Recognition with Attentive End-to-End Model ». In : *Proceedings of the 2020 International Conference on Multimodal Interaction*. 2020, p. 694-698.
- [42] Damien ERCEAU et Nicolas GUÉGUEN. « Tactile Contact and Evaluation of the Toucher ». In : *The Journal of Social Psychology* 147.4 (2007), 441-444. ISSN : 0022-4545. DOI : [10.3200/socp.147.4.441-444](https://doi.org/10.3200/socp.147.4.441-444).
- [43] Van ERP, Jan B. F et Alexander TOET. « Social Touch in Human-Computer Interaction ». English. In : *Frontiers in Digital Humanities* 2 (2015). ISSN : 2297-2668. DOI : [10.3389/fdigh.2015.00002](https://doi.org/10.3389/fdigh.2015.00002).
- [44] Angelica ESCALONA et al. « Brief report : improvements in the behavior of children with autism following massage therapy ». In : *Journal of autism and developmental disorders* 31.5 (2001), p. 513-516.
- [45] Ruth FELDMAN, Magi SINGER et Orna ZAGOORY. « Touch attenuates infants' physiological reactivity to stress ». In : *Developmental science* 13.2 (2010), p. 271-278.
- [46] Tiffany FIELD. « Touch for socioemotional and physical well-being : A review ». In : *Developmental Review* 30.4 (déc. 2010), p. 367-383. ISSN : 0273-2297. DOI : [10.1016/j.dr.2011.01.001](https://doi.org/10.1016/j.dr.2011.01.001).
- [47] Tiffany FIELD et al. « Massage therapy for infants of depressed mothers ». In : *Infant behavior and development* 19.1 (1996), p. 107-112.
- [48] Tiffany M FIELD et al. « Adolescents with attention deficit hyperactivity disorder benefit from massage therapy ». In : *Adolescence* 33.129 (1998), p. 103.
- [49] Jeffrey D. FISHER, Marvin RYTTING et Richard HESLIN. « Hands Touching Hands : Affective and Evaluative Effects of an Interpersonal Touch ». In : *Sociometry* 39.4 (1976), p. 416-421. ISSN : 0038-0431. DOI : [10.2307/3033506](https://doi.org/10.2307/3033506).
- [50] A. FLAGG et al. « Conductive fur sensing for a gesture-aware furry robot ». In : *2012 IEEE Haptics Symposium (HAPTICS)*. Mar. 2012, p. 99-104. DOI : [10.1109/HAPTIC.2012.6183776](https://doi.org/10.1109/HAPTIC.2012.6183776).
- [51] Anna FLAGG et Karon MACLEAN. « Affective touch gesture recognition for a furry zoomorphic machine ». en. In : ACM Press, 2013, p. 25. ISBN : 978-1-4503-1898-3. DOI : [10.1145/2460625.2460629](https://doi.org/10.1145/2460625.2460629).
- [52] Alberto GALLACE et Charles SPENCE. « The science of interpersonal touch : An overview ». In : *Neuroscience & Biobehavioral Reviews*. Touch,

- Temperature, Pain/Itch and Pleasure 34.2 (fév. 2010), p. 246-259. ISSN : 0149-7634. DOI : [10.1016/j.neubiorev.2008.10.004](https://doi.org/10.1016/j.neubiorev.2008.10.004).
- [53] Yona Falinie A. GAUS et al. « Social Touch Gesture Recognition Using Random Forest and Boosting on Distinct Feature Sets ». In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ICMI '15. New York, NY, USA : ACM, 2015, p. 399-406. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2830599](https://doi.org/10.1145/2818346.2830599).
- [54] Valeria GAZZOLA et al. « Primary somatosensory cortex discriminates affective significance in social touch ». en. In : *Proceedings of the National Academy of Sciences* 109.25 (juin 2012), E1657-E1666. ISSN : 0027-8424, 1091-6490. DOI : [10.1073/pnas.1113211109](https://doi.org/10.1073/pnas.1113211109).
- [55] Paul E GREENBAUM et Howard M ROSENFELD. « Varieties of touching in greetings : Sequential structure and sex-related differences ». In : *Journal of Nonverbal Behavior* 5.1 (1980), p. 13-25.
- [56] Laura K GUERRERO et Amy S EBESU. *While at Play : An Observational Analysis of Children's Touch During Interpersonal Interaction*. 1993.
- [57] Nicolas GUÉGUEN. « Nonverbal Encouragement of Participation in a Course : the Effect of Touching ». en. In : *Social Psychology of Education* 7.1 (mar. 2004), p. 89-98. ISSN : 1573-1928. DOI : [10.1023/B:SPOE.0000010691.30834.14](https://doi.org/10.1023/B:SPOE.0000010691.30834.14).
- [58] G.ZHAO et M.PIETIKAINEN. « Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.6 (2007), p. 915-928.
- [59] Antal HAANS et Wijnand IJSSELSTEIJN. « Mediated social touch : a review of current research and future directions ». en. In : *Virtual Reality* 9.2 (mar. 2006), p. 149-159. ISSN : 1434-9957. DOI : [10.1007/s10055-005-0014-2](https://doi.org/10.1007/s10055-005-0014-2).
- [60] Antal HAANS et Wijnand A IJSSELSTEIJN. « The virtual midas touch : helping behavior after a mediated social touch ». In : *IEEE Transactions on Haptics* 2.3 (2009), p. 136-140.
- [61] Markus HEINRICHS et al. « Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress ». eng. In : *Biological Psychiatry* 54.12 (déc. 2003), p. 1389-1398. ISSN : 0006-3223.
- [62] Susan S HENDRICK, Amy DICKE et Clyde HENDRICK. « The relationship assessment scale ». In : *Journal of social and personal relationships* 15.1 (1998), p. 137-142.

- [63] Maria HENRICSON et al. « The outcome of tactile touch on oxytocin in intensive care patients : a randomised controlled trial ». eng. In : *Journal of Clinical Nursing* 17.19 (oct. 2008), p. 2624-2633. ISSN : 1365-2702. DOI : [10.1111/j.1365-2702.2008.02324.x](https://doi.org/10.1111/j.1365-2702.2008.02324.x).
- [64] Alicia HERAZ et Manfred CLYNES. « Recognition of emotions conveyed by touch through force-sensitive screens : observational study of humans and machine learning techniques ». In : *JMIR mental health* 5.3 (2018), e10104.
- [65] Matthew J. HERTENSTEIN et Joseph J. CAMPOS. « Emotion Regulation Via Maternal Touch ». en. In : *Infancy* 2.4 (2001), p. 549-566. ISSN : 1532-7078. DOI : [10.1207/S15327078IN0204_09](https://doi.org/10.1207/S15327078IN0204_09).
- [66] Matthew J. HERTENSTEIN et al. « The communication of emotion via touch. » In : *Emotion* 9.4 (2009), p. 566-573. DOI : [10.1037/a0016108](https://doi.org/10.1037/a0016108).
- [67] Matthew J. HERTENSTEIN et al. « The communicative functions of touch in humans, nonhuman primates, and rats : a review and synthesis of the empirical research ». eng. In : *Genetic, Social, and General Psychology Monographs* 132.1 (fév. 2006), p. 5-94. ISSN : 8756-7547.
- [68] Matthew J. HERTENSTEIN et al. « Touch communicates distinct emotions. » In : *Emotion* 6.3 (2006), p. 528-533.
- [69] R HESLIN. *Steps toward a taxonomy of touching. Paper presented to the annual meeting of the Midwestern Psychological Association.* 1974.
- [70] Richard HESLIN, Tuan D. NGUYEN et Michele L. NGUYEN. « Meaning of touch : The case of touch from a stranger or same sex person ». en. In : *Journal of Nonverbal Behavior* 7.3 (mar. 1983), p. 147-157. ISSN : 1573-3653. DOI : [10.1007/BF00986945](https://doi.org/10.1007/BF00986945).
- [71] D. HUGHES, A. KRAUTHAMMER et N. CORRELL. « Recognizing social touch gestures using recurrent and convolutional neural networks ». In : *2017 IEEE International Conference on Robotics and Automation (ICRA)*. Mai 2017, p. 2315-2321. DOI : [10.1109/ICRA.2017.7989267](https://doi.org/10.1109/ICRA.2017.7989267).
- [72] Dana HUGHES et al. « Detecting and Identifying Tactile Gestures Using Deep Autoencoders, Geometric Moments and Gesture Level Features ». In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ICMI '15. New York, NY, USA : ACM, 2015, p. 415-422. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2830601](https://doi.org/10.1145/2818346.2830601).
- [73] G. HUISMAN. « Social Touch Technology : A Survey of Haptic Technology for Social Touch ». In : *IEEE Transactions on Haptics* 10.3 (juil. 2017), p. 391-408. ISSN : 1939-1412. DOI : [10.1109/TOH.2017.2650221](https://doi.org/10.1109/TOH.2017.2650221).

- [74] H. IWATA et S. SUGANO. « Human-robot-contact-state identification based on tactile recognition ». In : *IEEE Transactions on Industrial Electronics* 52.6 (déc. 2005), p. 1468-1477. ISSN : 0278-0046. DOI : [10.1109/TIE.2005.858739](https://doi.org/10.1109/TIE.2005.858739).
- [75] Z. JI et al. « Histogram based classification of tactile patterns on periodically distributed skin sensors for a humanoid robot ». In : *2011 RO-MAN*. Juil. 2011, p. 433-440. DOI : [10.1109/ROMAN.2011.6005261](https://doi.org/10.1109/ROMAN.2011.6005261).
- [76] Stanley E JONES. *The right touch : Understanding and using the language of physical contact*. Hampton Pr, 1994.
- [77] Stanley E. JONES et A. Elaine YARBROUGH. « A naturalistic study of the meanings of touch ». In : *Communication Monographs* 52.1 (mar. 1985), p. 19-56. ISSN : 0363-7751. DOI : [10.1080/03637758509376094](https://doi.org/10.1080/03637758509376094).
- [78] Sidney M. JOURARD. « An Exploratory Study of Body-Accessibility1 ». en. In : *British Journal of Social and Clinical Psychology* 5.3 (1966), p. 221-231. ISSN : 2044-8260. DOI : [10.1111/j.2044-8260.1966.tb00978.x](https://doi.org/10.1111/j.2044-8260.1966.tb00978.x).
- [79] Merel M JUNG. « Towards social touch intelligence : developing a robust system for automatic touch recognition ». In : *Proceedings of the 16th International Conference on Multimodal Interaction*. 2014, p. 344-348.
- [80] Merel M. JUNG et al. « Automatic recognition of touch gestures in the corpus of social touch ». en. In : *Journal on Multimodal User Interfaces* 11.1 (mar. 2017), p. 81-96. ISSN : 1783-7677, 1783-8738. DOI : [10.1007/s12193-016-0232-9](https://doi.org/10.1007/s12193-016-0232-9).
- [81] Merel M. JUNG et al. « Touch Challenge '15 : Recognizing Social Touch Gestures ». In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ICMI '15. New York, NY, USA : ACM, 2015, p. 387-390. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2829993](https://doi.org/10.1145/2818346.2829993).
- [82] Merel M. JUNG et al. « Touching the Void – Introducing CoST : Corpus of Social Touch ». In : *Proceedings of the 16th International Conference on Multimodal Interaction*. ICMI '14. New York, NY, USA : ACM, 2014, p. 120-127. ISBN : 978-1-4503-2885-2. DOI : [10.1145/2663204.2663242](https://doi.org/10.1145/2663204.2663242).
- [83] Takayuki KANDA et Hiroshi ISHIGURO. *Human-robot interaction in social robotics*. CRC Press, 2017.
- [84] Y. KIM et al. « A robust online touch pattern recognition for dynamic human-robot interaction ». In : *IEEE Transactions on Consumer Electronics* 56.3 (août 2010), p. 1979-1987. ISSN : 0098-3063. DOI : [10.1109/TCE.2010.5606355](https://doi.org/10.1109/TCE.2010.5606355).
- [85] Diederik P. KINGMA et Jimmy BA. *Adam : A Method for Stochastic Optimization*. 2014. arXiv : [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].

- [86] Louise P. KIRSCH et al. « Reading the mind in the touch : Neurophysiological specificity in the communication of emotions by touch ». en. In : *Neuropsychologia*. Special Issue : The Neuroscience of Empathy 116 (juil. 2018), p. 136-149. ISSN : 0028-3932. DOI : [10.1016/j.neuropsychologia.2017.05.024](https://doi.org/10.1016/j.neuropsychologia.2017.05.024).
- [87] H. KNIGHT et al. « Real-time social touch gesture recognition for sense robots ». In : *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2009, p. 3715-3720. DOI : [10.1109/IRoS.2009.5354169](https://doi.org/10.1109/IRoS.2009.5354169).
- [88] Kwan Min LEE et al. « Are physically embodied social agents better than disembodied social agents? : The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction ». In : *International Journal of Human-Computer Studies* 64.10 (2006), 962-973. ISSN : 1071-5819. DOI : [10.1016/j.ijhcs.2006.05.002](https://doi.org/10.1016/j.ijhcs.2006.05.002).
- [89] Haemy LEE MASSON et Hans Op de BEECK. « Socio-affective touch expression database ». en. In : *PLOS ONE* 13.1 (jan. 2018). Sous la dir. de Christos PAPADELIS, e0190921. ISSN : 1932-6203. DOI : [10.1371/journal.pone.0190921](https://doi.org/10.1371/journal.pone.0190921).
- [90] Julia A LEONARD, Talia BERKOWITZ et Anna SHUSTERMAN. « The effect of friendly touch on delay-of-gratification in preschool children ». In : *Quarterly Journal of Experimental Psychology* 67.11 (2014), p. 2123-2133.
- [91] Line S LÖKEN et al. « Coding of pleasant touch by unmyelinated afferents in humans ». In : *Nature neuroscience* 12.5 (2009), p. 547.
- [92] Kim MACLEAN. « The impact of institutionalization on child development ». In : *Development and psychopathology* 15.4 (2003), p. 853-884.
- [93] Ed MCDANIEL et Peter A. ANDERSEN. « None ». In : *Journal of Non-verbal Behavior* 22.1 (1998), 59-75. ISSN : 0191-5886. DOI : [10.1023/a:1022952509743](https://doi.org/10.1023/a:1022952509743).
- [94] Francis MCGLONE, Johan WESSBERG et Håkan OLAUSSON. « Discriminative and Affective Touch : Sensing and Feeling ». In : *Neuron* 82.4 (2014), 737-755. ISSN : 0896-6273. DOI : [10.1016/j.neuron.2014.05.001](https://doi.org/10.1016/j.neuron.2014.05.001).
- [95] Francis MCGLONE et al. « Discriminative touch and emotional touch. » In : *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 61.3 (2007), p. 173.

- [96] Sarah MCINTYRE et al. « Affective touch communication in close adult relationships ». In : *arXiv :1905.02613 [q-bio]* (mai 2019). arXiv : 1905.02613.
- [97] A. MONTAGU. *Touching : The Human Significance of the Skin*. Perennial library. HarperCollins, 1986. ISBN : 978-0-06-096028-5.
- [98] A. MONTAGU et T.L.A. MONTAGU. *Touching : The Human Significance of the Skin*. Harper & Row, 1978. ISBN : 9780060129798.
- [99] K. NAKAJIMA et al. « Emoballoon : A balloon-shaped interface recognizing social touch interactions ». In : *2013 IEEE Virtual Reality (VR)*. Mar. 2013, p. 1-4. DOI : [10.1109/VR.2013.6549433](https://doi.org/10.1109/VR.2013.6549433).
- [100] M. NARDELLI et al. « Classifying Affective Haptic Stimuli through Gender-Specific Heart Rate Variability Nonlinear Analysis ». In : *IEEE Transactions on Affective Computing* 11.3 (2020), p. 459-469.
- [101] F. NAYA, J. YAMATO et K. SHINOZAWA. « Recognizing human touching behaviors using a haptic interface for a pet-robot ». In : *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*. T. 2. Oct. 1999, 1030-1034 vol.2. DOI : [10.1109/ICSMC.1999.825404](https://doi.org/10.1109/ICSMC.1999.825404).
- [102] Joo-Hye PARK et al. « Touch Gesture Recognition System based on 1D Convolutional Neural Network with Two Touch Sensor Orientation Settings ». In : *2019 16th International Conference on Ubiquitous Robots (UR)*. IEEE. 2019, p. 65-70.
- [103] Miles L PATTERSON, Jack L POWELL et Mary G LENIHAN. « Touch, compliance, and interpersonal affect ». In : *Journal of Nonverbal behavior* 10.1 (1986), p. 41-50.
- [104] F. PEDREGOSA et al. « Scikit-learn : Machine Learning in Python ». In : *Journal of Machine Learning Research* 12 (2011), p. 2825-2830.
- [105] Leehe PELED-AVRON et al. « The role of empathy in the neural responses to observed human social touch ». eng. In : *Cognitive, Affective & Behavioral Neuroscience* 16.5 (2016), p. 802-813. ISSN : 1531-135X. DOI : [10.3758/s13415-016-0432-5](https://doi.org/10.3758/s13415-016-0432-5).
- [106] Rosalind W. PICARD. « Affective computing : challenges ». en. In : *International Journal of Human-Computer Studies* 59.1-2 (juil. 2003), p. 55-64. ISSN : 10715819. DOI : [10.1016/S1071-5819\(03\)00052-1](https://doi.org/10.1016/S1071-5819(03)00052-1).
- [107] R.W. PICARD, E. VYZAS et J. HEALEY. « Toward machine emotional intelligence : analysis of affective physiological state ». en. In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.10 (oct. 2001), p. 1175-1191. ISSN : 01628828. DOI : [10.1109/34.954607](https://doi.org/10.1109/34.954607).

- [108] Odile PLAISANT et al. « Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R ». In : *Annales Médico-psychologiques, revue psychiatrique*. T. 168. 2. Elsevier. 2010, p. 97-106.
- [109] M. S. REMLAND, T. S. JONES et H. BRINKMAN. « Interpersonal distance, body orientation, and touch : effects of culture, gender, and age ». eng. In : *The Journal of Social Psychology* 135.3 (juin 1995), p. 281-297. ISSN : 0022-4545. DOI : [10.1080/00224545.1995.9713958](https://doi.org/10.1080/00224545.1995.9713958).
- [110] Martin S. REMLAND, Tricia S. JONES et Heidi BRINKMAN. « Proxemic and haptic behavior in three European countries ». In : *Journal of Non-verbal Behavior* 15.4 (1991), 215–232. ISSN : 0191-5886. DOI : [10.1007/bf00986923](https://doi.org/10.1007/bf00986923).
- [111] Eleanor H ROSCH. « On the internal structure of perceptual and semantic categories ». In : *Cognitive development and acquisition of language*. Elsevier, 1973, p. 111-144.
- [112] Ognjen RUDOVIC et al. « Personalized machine learning for robot perception of affect and engagement in autism therapy ». In : *Science Robotics* 3.19 (2018).
- [113] James A. RUSSELL. « A circumplex model of affect ». In : *Journal of Personality and Social Psychology* 39.6 (1980), p. 1161-1178. ISSN : 1939-1315(Electronic),0022-3514(Print). DOI : [10.1037/h0077714](https://doi.org/10.1037/h0077714).
- [114] James A RUSSELL, Anna WEISS et Gerald A MENDELSON. « Affect grid : a single-item scale of pleasure and arousal. » In : *Journal of personality and social psychology* 57.3 (1989), p. 493.
- [115] Shafiq Ur RÉHMAN et Li LIU. « iFeeling : Vibrotactile Rendering of Human Emotions on Mobile Phones ». In : *Information Security and Privacy*. Information Security et Privacy, 2010, 1–20. DOI : [10.1007/978-3-642-12349-8_1](https://doi.org/10.1007/978-3-642-12349-8_1).
- [116] Jyotirmay SANGHVI et al. « Automatic analysis of affective postures and body motion to detect engagement with a game companion ». In : *Proceedings of the 6th international conference on Human-robot interaction*. 2011, p. 305-312.
- [117] Coralie SANN et Arlette STRERI. « The limits of newborn’s grasping to detect texture in a cross-modal transfer task ». In : *Infant Behavior and Development* 31.3 (2008), 523–531. ISSN : 0163-6383. DOI : [10.1016/j.infbeh.2008.03.001](https://doi.org/10.1016/j.infbeh.2008.03.001).

- [118] Alexandre SCHAEFER et al. « Assessing the effectiveness of a large database of emotion-eliciting films : A new tool for emotion researchers ». In : *Cognition and Emotion* 24.7 (nov. 2010), p. 1153-1172. ISSN : 0269-9931. DOI : [10.1080/02699930903274322](https://doi.org/10.1080/02699930903274322).
- [119] Michael SCHAEFER et al. « Mirror-like brain responses to observed touch and personality dimensions ». English. In : *Frontiers in Human Neuroscience* 7 (2013). ISSN : 1662-5161. DOI : [10.3389/fnhum.2013.00227](https://doi.org/10.3389/fnhum.2013.00227).
- [120] Klaus R. SCHERER. « Studying the emotion-antecedent appraisal process : An expert system approach ». In : *Cognition and Emotion* 7.3-4 (1993), 325–355. ISSN : 0269-9931. DOI : [10.1080/02699939308409192](https://doi.org/10.1080/02699939308409192).
- [121] Takanori SHIBATA, Yukitaka KAWAGUCHI et Kazuyoshi WADA. « Investigation on People Living with Seal Robot at Home ». en. In : *International Journal of Social Robotics* 4.1 (jan. 2012), p. 53-63. ISSN : 1875-4805. DOI : [10.1007/s12369-011-0111-1](https://doi.org/10.1007/s12369-011-0111-1).
- [122] Robert SHUTER. « Proxemics and tactility in Latin America ». In : *Journal of communication* 26.3 (1976), p. 46-52.
- [123] David SILVERA-TAWIL, David RYE et Mari VELONAKI. « Artificial skin and tactile sensing for socially interactive robots : A review ». In : *Robotics and Autonomous Systems. Advances in Tactile Sensing and Touch-based Human Robot Interaction* 63 (jan. 2015), p. 230-243. ISSN : 0921-8890. DOI : [10.1016/j.robot.2014.09.008](https://doi.org/10.1016/j.robot.2014.09.008).
- [124] David SILVERA-TAWIL, David RYE et Mari VELONAKI. « Interpretation of Social Touch on an Artificial Arm Covered with an EIT-based Sensitive Skin ». en. In : *International Journal of Social Robotics* 6.4 (nov. 2014), p. 489-505. ISSN : 1875-4805. DOI : [10.1007/s12369-013-0223-x](https://doi.org/10.1007/s12369-013-0223-x).
- [125] David SILVERA TAWIL, David RYE et Mari VELONAKI. « Interpretation of the modality of touch on an artificial arm covered with an EIT-based sensitive skin ». en. In : *The International Journal of Robotics Research* 31.13 (nov. 2012), p. 1627-1641. ISSN : 0278-3649. DOI : [10.1177/0278364912455441](https://doi.org/10.1177/0278364912455441).
- [126] Harsimrat SINGH et al. « The brain's response to pleasant touch : an EEG investigation of tactile caressing ». In : *Frontiers in Human Neuroscience* 8 (2014). ISSN : 1662-5161. DOI : [10.3389/fnhum.2014.00893](https://doi.org/10.3389/fnhum.2014.00893).
- [127] David E SMITH, Frank N WILLIS et Joseph A GIER. « Success and interpersonal touch in a competitive setting ». In : *Journal of nonverbal behavior* 5.1 (1980), p. 26-34.

- [128] Mandayam A SRINIVASAN. « What is haptics ? » In : *Laboratory for Human and Machine Haptics : The Touch Lab, Massachusetts Institute of Technology* (1995), p. 1-11.
- [129] Dale M STACK et Darwin W MUIR. « Tactile stimulation as a component of social interchange : New interpretations for the still-face effect ». In : *British Journal of Developmental Psychology* 8.2 (1990), p. 131-145.
- [130] W. D. STIEHL et al. « Design of a therapeutic robotic companion for relational, affective touch ». In : *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. Août 2005, p. 408-415. DOI : [10.1109/ROMAN.2005.1513813](https://doi.org/10.1109/ROMAN.2005.1513813).
- [131] Jiong SUN et al. « Tactile Interaction and Social Touch : Classifying Human Touch Using a Soft Tactile Sensor ». en. In : ACM Press, 2017, p. 523-526. ISBN : 978-1-4503-5113-3. DOI : [10.1145/3125739.3132614](https://doi.org/10.1145/3125739.3132614).
- [132] Juulia T. SUVILEHTO. *Maintaining social bonds via touching : A cross-cultural study*. en. Aalto University, 2018. ISBN : 978-952-60-8264-6.
- [133] Viet Cuong TA et al. « The Grenoble System for the Social Touch Challenge at ICMI 2015 ». en. In : nov. 2015. DOI : [10.1145/2818346.2830598](https://doi.org/10.1145/2818346.2830598).
- [134] Viet-Cuong TA et al. « The Grenoble System for the Social Touch Challenge at ICMI 2015 ». In : *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ICMI '15*. New York, NY, USA : ACM, 2015, p. 391-398. ISBN : 978-1-4503-3912-4. DOI : [10.1145/2818346.2830598](https://doi.org/10.1145/2818346.2830598).
- [135] Kohske TAKAHASHI et al. « Improving shared experiences by haptic telecommunication ». In : *2011 International Conference on Biometrics and Kansei Engineering*. IEEE. 2011, p. 210-215.
- [136] Erin H THOMPSON et James A HAMPTON. « The effect of relationship status on communicating emotions through touch ». In : *Cognition and emotion* 25.2 (2011), p. 295-306.
- [137] Chantal TRISCOLI et al. « Touch between romantic partners : Being stroked is more pleasant than stroking and decelerates heart rate ». In : *Physiology & behavior* 177 (2017), p. 169-175.
- [138] Mohamed Yassine TSALAMLAL et al. « Haptic communication of dimensions of emotions using air jet based tactile stimulation ». In : *Journal on Multimodal User Interfaces* 9.1 (2015), p. 69-77.
- [139] Ashish VASWANI et al. « Attention is All You Need ». In : 2017.

- [140] Rongrong WANG et al. « Keep in touch : channel, expectation and experience ». In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, p. 139-148.
- [141] Sandra J WEISS. « Measurement of the sensory qualities in tactile interaction. » In : *Nursing Research* (1992).
- [142] Frank H WILHELM et al. « Social anxiety and response to touch : incongruence between self-evaluative and physiological reactions ». In : *Biological Psychology* 58.3 (déc. 2001), p. 181-202. ISSN : 0301-0511. DOI : [10.1016/S0301-0511\(01\)00113-2](https://doi.org/10.1016/S0301-0511(01)00113-2).
- [143] Sharon J WILLIAMS et Frank N WILLIS. « Interpersonal touch among preschool children at play ». In : *The Psychological Record* 28.4 (1978), p. 501-508.
- [144] Frank N WILLIS et Gale E HOFMANN. « Development of tactile patterns in relation to age, sex, and race. » In : *Developmental Psychology* 11.6 (1975), p. 866.
- [145] Frank N WILLIS et Dennis L REEVES. « Touch interactions in junior high students in relation to sex and race. » In : *Developmental Psychology* 12.1 (1976), p. 91.
- [146] Siewart van WINGERDEN et al. « A Neural Network Based Approach to Social Touch Classification ». In : *Proceedings of the 2014 Workshop on Emotion Representation and Modelling in Human-Computer-Interaction-Systems*. ERM4HCI '14. New York, NY, USA : ACM, 2014, p. 7-12. ISBN : 978-1-4503-0124-4. DOI : [10.1145/2668056.2668060](https://doi.org/10.1145/2668056.2668060).
- [147] Richard K WINKELMANN. « The erogenous zones : their nerve supply and its significance. » In : *Proceedings of the staff meetings. Mayo Clinic*. T. 34. 2. 1959, p. 39.
- [148] Zichao YANG et al. « Hierarchical Attention Networks for Document Classification ». In : *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*. San Diego, California : Association for Computational Linguistics, juin 2016, p. 1480-1489. DOI : [10.18653/v1/N16-1174](https://doi.org/10.18653/v1/N16-1174).
- [149] Steve YOHANAN et Karon E. MACLEAN. « The Role of Affective Touch in Human-Robot Interaction : Human Intent and Expectations in Touching the Haptic Creature ». en. In : *International Journal of Social Robotics* 4.2 (avr. 2012), p. 163-180. ISSN : 1875-4805. DOI : [10.1007/s12369-011-0126-7](https://doi.org/10.1007/s12369-011-0126-7).

- [150] Steven John YOHANAN. « The Haptic Creature : social human-robot interaction through affective touch ». eng. Thèse de doct. University of British Columbia, 2012. DOI : [10.14288/1.0052135](https://doi.org/10.14288/1.0052135).
- [151] Bo ZHOU et al. « Textile pressure mapping sensor for emotional touch detection in human-robot interaction ». In : *Sensors* 17.11 (2017), p. 2585.
- [152] Nan ZHOU et Jun DU. « Recognition of Social Touch Gestures Using 3D Convolutional Neural Networks ». en. In : *Pattern Recognition*. Sous la dir. de Tieniu TAN et al. Communications in Computer and Information Science. Singapore : Springer, 2016, p. 164-173. ISBN : 978-981-10-3002-4. DOI : [10.1007/978-981-10-3002-4_14](https://doi.org/10.1007/978-981-10-3002-4_14).

Annexe A

Questionnaires

A.1 Big Five Inventory français (BFI-Fr)

Instructions :

Pour chaque item, indiquez dans quelle mesure la phrase vous correspond. LISEZ CHAQUE ITEM ATTENTIVEMENT AVANT DE RÉPONDRE. Répondez le plus honnêtement possible. Merci.

1. désapprouve fortement
2. désapprouve un peu
3. n'approuve ni ne désapprouve
4. approuve un peu
5. approuve fortement

Je me vois comme quelqu'un qui...

1. _ est bavard
2. _ a tendance à critiquer les autres
3. _ travaille consciencieusement
4. _ est déprimé, cafardeux
5. _ créatif, plein d'idées originales
6. _ est réservé
7. _ est serviable est n'est pas égoïste avec les autres
8. _ peut être parfois négligent
9. _ est "relax", détendu, gère bien le stress
10. _ s'intéresse à de nombreux sujets
11. _ est plein d'énergie
12. _ commence facilement à se disputer avec les autres
13. _ est fiable dans son travail

14. _ peut être angoissé
15. _ est ingénieux
16. _ communique beaucoup d'enthousiasme
17. _ est indulgent de nature
18. _ a tendance à être désorganisé
19. _ se tourmente beaucoup
20. _ a une grande imagination
21. _ a tendance à être silencieux
22. _ fait généralement confiance aux autres
23. _ a tendance à être paresseux
24. _ comme quelqu'un de tempéré, pas facilement troublé
25. _ est inventif
26. _ a une forte personnalité, s'exprime avec assurance
27. _ est parfois dédaigneux, méprisant
28. _ persévère jusqu'à ce que sa tâche soit finie
29. _ peut être lunatique d'humeur changeante
30. _ apprécie les activités artistiques et esthétiques
31. _ est quelquefois timide, inhibé
32. _ est prévenant et gentil avec presque tout le monde
33. _ est efficace dans son travail
34. _ reste calme dans les situations angoissantes
35. _ préfère un travail simple et routinier
36. _ est sociable, extraverti
37. _ est parfois impoli avec les autres
38. _ fait des projets et les poursuit
39. _ est facilement anxieux
40. _ aime réfléchir et jouer avec des idées
41. _ est peu intéressé par tout ce qui est artistique
42. _ aime coopérer avec les autres
43. _ est facilement distrait
44. _ a de bonne connaissance en art, musique ou en littérature
45. _ cherche des histoires aux autres

Score= moyenne de items dans la dimension

E (Extraversion, Energie, Enthousiasme)

8 items : 1, 6R, 11, 16, 21R, 26, 31R, 36;

A (Agréabilité, Altruisme, Affection)

10 items : 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42, 45R;

A.2. Index de Réactivité Interpersonnel (IRI-FR)

C (Conscience, Contrôle, Contrainte)

9 items : 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R;

N (émotion Négatives, Névrosisme, Nevrosité)

8 items : 4, 9R, 14, 19, 24R, 29, 34R, 39;

O (Ouverture, Originalité, Ouverture d'esprit)

10 items : 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44;

Chaque facteur correspond à la moyenne de la somme des items en inversant les items marqués d'un R (Reverse) (5 devenant 1; 4 2, 3 3....)

BFI	Original study [108]	
N	738 Males	1654 Females
E	3.1 (0.7)	3.2 (0.7)
A	3.8 (0.5)	3.9 (0.5)
C	3.1 (0.6)	3.5 (0.6)
N	2.6 (0.7)	3.1 (0.8)
O	3.1 (0.6)	3.3 (0.6)

TABLE A.1 – Mean and std of personality traits

A.2 Index de Réactivité Interpersonnel (IRI-FR)

Instructions :

Le questionnaire suivant concerne vos pensées et vos sentiments dans différentes situations. Pour chaque item, indiquez dans quelle mesure la phrase vous correspond en choisissant la lettre appropriée sur l'échelle décrite en haut de cette page : A, B, C, D, ou E. Quand vous avez choisi votre réponse, cochez la lettre sur la feuille de réponse à côté du numéro de l'item. LISEZ CHAQUE ITEM ATTENTIVEMENT AVANT DE RÉPONDRE. Répondez le plus honnêtement possible. Merci.

- A Ne me décrit pas bien
- B Me décrit un peu
- C Me décrit moyennement
- D Me décrit assez bien
- E Me décrit très bien

1. J'imagine et fantasme assez régulièrement sur des choses qui pourraient m'arriver.(FS)
2. Je suis souvent sensible à l'égard des gens qui ont moins de chance que moi et me préoccupe de leur sort.(EC)
3. J'ai parfois des difficultés à imaginer les choses du point de vue d'une autre personne.(PT) (-)
4. Parfois, je ne me sens pas vraiment désolé pour les autres lorsqu'ils ont des problèmes.(EC) (-)
5. Je m'identifie complètement par les sentiments aux personnages d'un roman.(FS)
6. Dans les situations d'urgences, je me sens inquiet et mal à l'aise.(PD)
7. Je suis habituellement objectif (ve) lorsque je regarde un film ou une pièce de théâtre, et je ne me laisse pas souvent captiver entièrement dans l'action.(FS) (-)
8. En cas de désaccord, j'essaie de considérer le point de vue de chacun avant de prendre une décision.(PT)
9. Lorsque je vois que l'on profite de quelqu'un, j'éprouve une certaine envie de le protéger. (EC)
10. Quand je me retrouve dans une situation très émouvante, je me sens parfois incapable de réagir. (PD)
11. J'essaie parfois de mieux comprendre mes amis en imaginant comment les choses se présentent de leur point de vue. (PT)
12. Se sentir totalement impliqué dans un bon livre ou un bon film est assez rare pour moi.(FS) (-)
13. Lorsque je vois quelqu'un souffrir, j'ai tendance à garder mon sang froid.(PD) (-)
14. Habituellement, le malheur des autres ne me perturbe pas particulièrement. (EC) (-)
15. Si je suis sûr que j'ai raison, je ne perds pas mon temps à écouter les arguments des autres.(PT) (-)
16. Après avoir vu une pièce de théâtre ou un film, il m'est arrivé de me sentir dans la peau d'un des personnages. (FS)
17. Je suis effrayé à l'idée de me retrouver dans une situation émotionnellement tendue. (PD)
18. Lorsque je vois une personne traitée injustement, il m'arrive parfois de ne pas ressentir beaucoup de pitié pour elle.(EC) (-)
19. Je suis habituellement efficace dans la gestion des situations d'urgences.(PD) (-)
20. Je suis souvent touché (e) par des événements dont je suis témoin. (EC)

A.2. *Index de Réactivité Interpersonnel (IRI-FR)*

21. Je crois qu'il y a deux facettes à chaque question et j'essaie de prendre en considération toutes les deux.(PT)
22. Je me décrirais comme une personne qui s'attendrit plutôt facilement.(EC)
23. Lorsque je regarde un bon film, je peux très facilement me mettre à la place d'un des personnages principaux. (FS)
24. J'ai tendance à perdre mon sang-froid dans les situations d'urgence.(PD)
25. Quand je suis en colère contre quelqu'un, j'essaye généralement de me mettre à sa place un instant.(PT)
26. Lorsque je lis un roman (ou une histoire) intéressant, j'imagine comment je me sentirais si les événements de l'histoire m'arrivaient.(FS)
27. Quand je vois quelqu'un qui a vraiment besoin d'aide dans une situation d'urgence, je perds mes moyens.(PD)
28. Avant de critiquer quelqu'un, j'essaie d'imaginer ce que je ressentirais si j'étais à sa place.(PT)

NOTE :(-) denotes item to be scored in reverse fashion

PT= perspective-taking scale

FS= fantasy scale

EC= empathic concern scale

PD= personal distress scale

A=0

B=1

C=2

D=3

E=4

Except for reversed-scored items, which are scored :

A=4

B=3

C=2

D=1

E=0

Scores are obtained for each sub-scale by summing the respective items score.

Empathy	Original study [34]	
N	579 Males	582 Females
Fantasy (FS)	15.73 (5.60)	18.75 (5.17)
Perspective Taking (PT)	16.78 (4.72)	17.96 (4.85)
Empathic Concern (EC)	19.04 (4.21)	21.67 (3.83)
Personal Distress (PD)	9.46 (4.55)	12.28 (5.01)

TABLE A.2 – Mean and std of Empathy

Annexe B

Valence and Arousal recognition

B.1 Arousal and Valence recognition

B.1.1 Arousal recognition

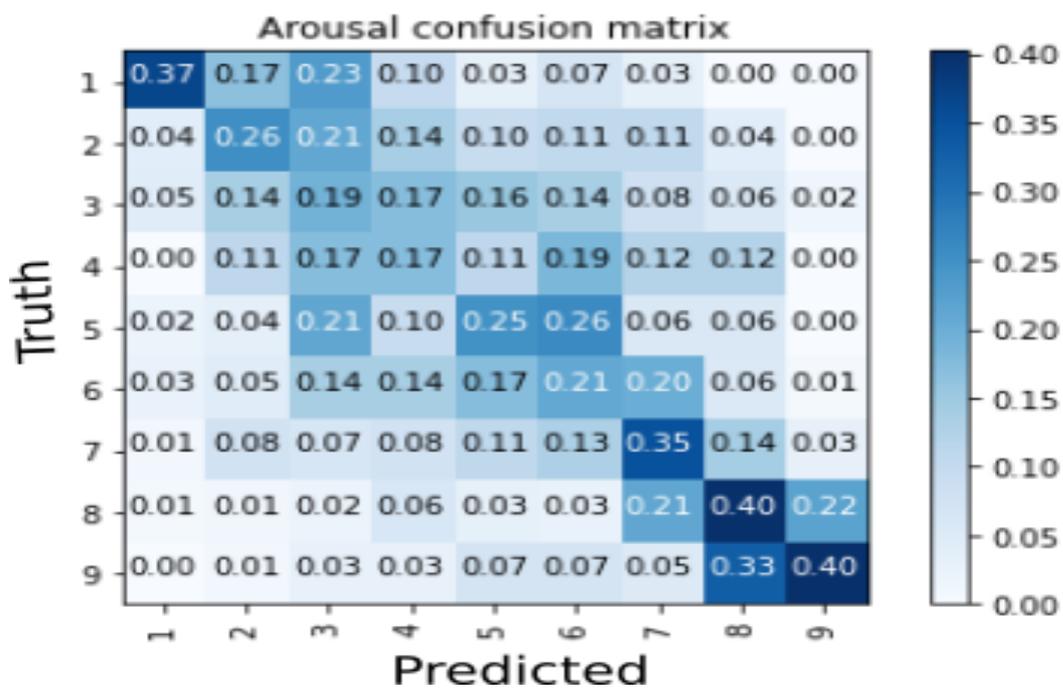


FIGURE B.1 – Arousal confusion matrix.

Arousal	X	X+IT	stratified	most_frequent	uniform
1	0.2(0.23)	0.33(0.22)	0.04(0.08)	0.0(0.0)	0.03(0.05)
2	0.22(0.13)	0.28(0.1)	0.04(0.05)	0.0(0.0)	0.09(0.07)
3	0.21(0.03)	0.2(0.04)	0.16(0.03)	0.0(0.0)	0.12(0.08)
4	0.15(0.03)	0.16(0.05)	0.08(0.06)	0.0(0.0)	0.14(0.08)
5	0.17(0.02)	0.23(0.04)	0.11(0.12)	0.0(0.0)	0.15(0.09)
6	0.16(0.08)	0.2(0.11)	0.11(0.06)	0.0(0.0)	0.17(0.05)
7	0.22(0.08)	0.29(0.08)	0.13(0.04)	0.0(0.0)	0.16(0.06)
8	0.4(0.06)	0.42(0.03)	0.11(0.04)	0.28(0.0)	0.15(0.06)
9	0.39(0.03)	0.45(0.03)	0.08(0.05)	0.0(0.0)	0.14(0.05)
accuracy	0.25(0.03)	0.29(0.03)	0.11(0.02)	0.16(0.0)	0.13(0.03)
macro avg	0.24(0.05)	0.28(0.04)	0.09(0.02)	0.03(0.0)	0.13(0.03)

TABLE B.1 – Arousal classification results for RF_X , RF_{X+IT} and baselines. The metric used are f1-scores, f1-scores average and accuracy.

B.1.2 Valence recognition

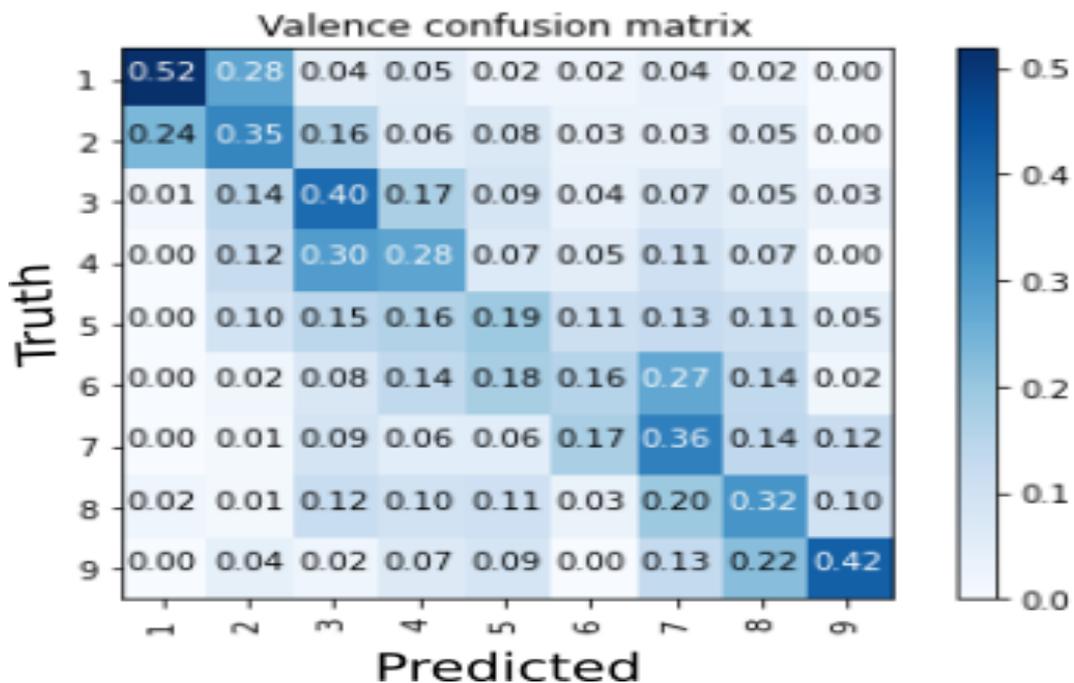


FIGURE B.2 – Valence confusion matrix.

Valence	X	X+IT	stratified	most_frequent	uniform
1	0.54(0.1)	0.61(0.08)	0.16(0.06)	0.0(0.0)	0.07(0.04)
2	0.28(0.1)	0.33(0.13)	0.16(0.06)	0.0(0.0)	0.09(0.03)
3	0.36(0.01)	0.4(0.06)	0.11(0.05)	0.28(0.0)	0.11(0.04)
4	0.14(0.06)	0.22(0.05)	0.13(0.02)	0.0(0.0)	0.13(0.07)
5	0.12(0.07)	0.17(0.12)	0.12(0.05)	0.0(0.0)	0.09(0.04)
6	0.1(0.08)	0.13(0.12)	0.06(0.06)	0.0(0.0)	0.05(0.07)
7	0.24(0.04)	0.36(0.07)	0.13(0.04)	0.0(0.0)	0.08(0.07)
8	0.29(0.06)	0.32(0.08)	0.2(0.1)	0.0(0.0)	0.07(0.07)
9	0.3(0.11)	0.38(0.11)	0.05(0.04)	0.0(0.0)	0.06(0.04)
accuracy	0.3(0.02)	0.36(0.03)	0.13(0.01)	0.16(0.0)	0.09(0.02)
macro avg	0.26(0.01)	0.33(0.03)	0.13(0.01)	0.03(0.0)	0.08(0.03)

TABLE B.2 – Valence classification results for RF_X , RF_{X+IT} and baselines. The metric used are f1-scores, f1-scores average and accuracy.