Towards An Emotionally Communicative Robot:

Feature Analysis for Multimodal Support of Affective Touch Recognition

by

Xi Laura Cang

BSc Mathematics, The University of British Columbia, 2007 BEd Secondary, The University of British Columbia, 2010 BCS, The University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Computer Science)

The University of British Columbia (Vancouver)

August 2016

© Xi Laura Cang, 2016

Abstract

Human affective state extracted from touch interaction takes advantage of natural communication of emotion through physical contact, enabling applications like robot therapy [88, 104], intelligent tutoring systems [25], emotionally-reactive smart tech, and more. This work focused on the emotionally aware robot pet context and produced a custom, low-cost piezoresistive fabric touch sensor at 1-inch taxel resolution that accommodates the flex and stretch of the robot in motion. Using established machine learning techniques, we built classification models of social and emotional touch data. We present an iteration of the human-robot interaction loop for an emotionally aware robot [110] through two distinct studies and demonstrate gesture recognition at roughly 85% accuracy (chance 14%).

The first study collected social touch gesture data (N=26) to assess data quality of our custom sensor under noisy conditions: mounted on a robot skeleton simulating regular breathing, obscured under fur casings, placed over deformable surfaces.

Our second study targeted affect with the same sensor, wherein participants (N=30) relived emotionally intense memories while interacting with a smaller stationary robot, generating touch data imbued with the following: *Stressed*, *Excited*, *Relaxed*, or *Depressed*. A feature space analysis triangulating touch, gaze, and physiological data highlighted the dimensions of touch that suggest affective state.

To close the interactive loop, we had participants (N=20) evaluate researcherdesigned breathing behaviours on 1-DOF robots for emotional content. Results demonstrate that these behaviours can display human-recognizable emotion as perceptual affective qualities across the valence-arousal emotion model [83]. Finally, we discuss the potential impact of a system capable of emotional "conversation" [89] with human users, referencing specific applications.

Preface

I have been fortunate to have been able to collaborate with a number of people throughout the studies described in Chapters 2 - 4. For each work, I describe the nature of my role and recognize my colleagues for their contribution.

Chapter 2 (Gesture Classification) is a conference paper published at International Conference on Multimodal Interaction (ICMI 2015)¹ on touch gesture classification and is presented here in its entirety. My co-authors include labmates Paul Bucci (fellow MSc student), Andrew Strang (Research Engineer), Jeff Allen (PhD student), Sean Liu (summer undergraduate research assistant), and supervisor Dr. Karon MacLean. The study concept, analysis procedures, and full paper writing were done by me. However, collaborators provided regular feedback and helped refine my original study proposal, informing the resulting study design and analysis process. Furthermore, Paul and Sean worked with me to conduct the actual data collection for the latter half of the study. I carried out final editing in conjunction with Paul and Karon.

A version of the paper intended for an upcoming journal submission forms Chapter 3 (Affect Detection). The featured study describes the collection and classification emotion data from gaze and biometric support of touch data. This is again a collaborative effort between our Sensory Perception and Interaction (SPIN) lab and Dr. Jussi Rantala, a post-doc visitng from the University of Tampere in Finland at the time of the study. The author list consists of myself, Dr. Jussi Rantala, labmate Paul Bucci, and supervisor Dr. Karon MacLean. My independent contri-

¹Cang XL, Bucci P, Strang A, Allen J, MacLean K, Liu HY. 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 2015 Nov 9 (pp. 147-154). ACM.

bution to this project includes the original study proposal, all touch-related background research and materials for data collection, the integrated multimodal analysis, and the final paper writing. Again, our team met regularly to collaboratively refine and improve the study design, enlisting the expertise of Dr Jessica Tracy of UBC's Emotion Lab before deciding on an emotion elicitation methodology. Jussi and I wrote the logging software for integration of gaze and touch data together; he brought gaze expertise, and I touch. The two of us then ran participants to collect multimodal emotion data. During the analysis phase, each of Jussi, Paul, and I were responsible for data cleaning of the three data sets: gaze, biometric, and touch respectively. We met weekly via Skype to discuss data format, cleaning techniques, and final feature set calculations, sharing our findings and data sets. Classification scripts to generate results were pair programmed by myself and Paul. I drafted the initial manuscript, although the entire team was involved in multiple editing passes. Chapter 3 includes all of our results and findings; while not the exact final journal text, it will bear a close resemblance.

The robot behaviour recognition of Chapter 4 (Behaviour Sketching) is an early study that has informed current work on an expanded set of robot behaviour generation and interpretation. The content here has not appeared elsewhere and will be included in a larger work, as the first of a sequence of design studies. It was also a collaboration between myself, visiting PhD student Merel Jung of the University of Twente in the Netherlands, Dr. Jussi Rantala, and Paul Bucci. Study design refinement was a full group effort built on ideas generated by Merel and myself. Behaviour creation and participant materials were designed by myself, Jussi, and Merel with final study interface coding done in pair programming fashion by myself and Jussi. Robot hardware is attributed to Paul Bucci alone. Data analysis was planned and conducted by myself and Merel, and while I took lead on writing, edits were made iteratively by the whole team.

This work was approved by the University of British Columbia (UBC) Behavioural Research Ethics Board (BREB) with ethics number #H15-02611. Study forms can be found in Appendix A.

Finally, I take responsibility for the concept of this thesis, chapter integration, and all other formal writing, including any and all errata.

Table of Contents

Ab	strac	t	•••••	•••	••	••	• •	•	•	ii
Pro	eface	• • • •		•••	••	••	• •	•	•	iii
Tal	ble of	Conter	ts	•••	••	••	••	•	•	v
Lis	st of T	ables .		•••	••	••	• •	•	•	viii
Lis	st of F	ligures		•••	••	••	• •	•	•	X
Gle	ossar	y		•••	••	••	• •	•	•	xiv
Ac	know	ledgme	nts	•••	••	••	• •	•	•	XV
De	dicati	ion		•••	••	••	• •	•	•	xvi
1	Intro	oductio			••			•	•	1
	1.1	Backgi	ound							3
		1.1.1	Robot platforms							3
		1.1.2	The CuddleBot and CuddleBits							7
		1.1.3	Sensing technology and classification te	chni	ique	s.				7
		1.1.4	Interaction Styles							8
	1.2	Approa	ch							10
	1.3	Thesis	Organization					•		11
2	Gest	ure Cla	ssification	•••	••	••	• •	•	•	13

	2.1	Introd	uction	14
		2.1.1	Questions and Contributions	15
		2.1.2	Applications	16
		2.1.3	Detailed Requirements	17
	2.2	Relate	d Work	18
		2.2.1	Social & Affective Touch Communication	18
		2.2.2	Flexible Pressure-Location Sensors	19
	2.3	Studie	8	20
		2.3.1	Apparatus	20
		2.3.2	Methods	21
		2.3.3	Analysis and Results	23
	2.4	Discus	ssion	29
	2.5	Conclu	usions	33
	2.6	Future	Work	34
	2.7	Ackno	wledgments	35
3	Affe	ct Dete	ction	36
	3.1	Introd	uction	38
		3.1.1	Approach and Research Questions	40
		3.1.2	Contributions	42
	3.2	Relate	d Work	43
		3.2.1	Emotion Set	43
		3.2.2	Emotion Elicitation	44
		3.2.3	Modalities	45
	3.3	Metho	bds	47
		3.3.1	Experimental Setup	47
		3.3.2	Procedure	51
	3.4	Result	8	61
		3.4.1	Pressure-location Feature Set: Emotion Classification	61
		3.4.2	Frequency Domain Feature Set: Emotion Classification	64
		3.4.3	Participant Classification	67
		3.4.4	Feature Set Analysis	67
		3.4.5	Experienced Emotion Trajectory and Interview Results	70

5.5	Discussion	74
	3.5.1 Findings	74
	3.5.2 Experimental Methodology	75
	3.5.3 Application Implications and Future Work	77
3.6	Conclusion	79
Beh	aviour Sketching	81
4.1	Related Work	82
4.2	Experiment Method	83
4.3	Results	85
4.4	Discussion	88
Con	clusions and Future Work	91
51	Outcomes and Impacts	92
5.1	1	/ -
5.2	Future Work	93
5.2 bliog	Future Work	93 95
5.2 5.2 bliogi Supj	Future Work	93 95 107
5.2 bliogr Supj A.1	Future Work raphy porting Materials Study Forms	 93 95 107 107
5.2 bliogr Supj A.1	Future Work	 93 95 107 107 107
5.2 bliogr Sup A.1	Future Work Future Work raphy Future Work porting Materials Future Work Study Forms Future Work A.1.1 Consent Form A.1.2 Call for Participation Form	 93 95 107 107 107 110
5.2 bliogr Sup A.1 A.2	Future Work Future Work raphy Future Work porting Materials Future Work Study Forms Future Work A.1.1 Consent Form A.1.2 Call for Participation Form Participant Response Forms Future Work	 93 95 107 107 110 112
5.2 bliogr Sup A.1 A.2	Future Work	93 93 95 107 107 107 110 112 112 112
5.2 bliogr A.1 A.2	Future Work Future Work raphy Future Work porting Materials Future Work porting Materials Future Work Study Forms Future Work A.1.1 Consent Form A.1.2 Call for Participation Form Participant Response Forms Future Work A.2.1 Gesture Study Demographic Questionnaire A.2.2 Affective Rating Form	93 93 95 107 107 107 110 112 112 112 115
	3.6 Beha 4.1 4.2 4.3 4.4 Con	3.5.1 Findings 3.5.2 Experimental Methodology 3.5.3 Application Implications and Future Work 3.6 Conclusion 4.1 Related Work 4.2 Experiment Method 4.3 Results 4.4 Discussion 5.1 Outcomes and Impacts

List of Tables

Table 1.1	An overview of robot platforms intended for therapeutic use	5
Table 2.1	Touch gesture instructions as provided to participants	22
Table 3.1	Experimental Procedure Summary	52
Table 3.2	Summary of the condition sets for Neutralization and Emotion	
	tasks. For example, one participant session consisted of R1 neu-	
	tral task, Stressed emotion task, R2 neutral task, and Relaxed	
	emotion task	54
Table 3.3	Touch gesture instructions as provided to participants [19]	54
Table 3.4	A motivating overview of analysis techniques.	58
Table 3.5	A motivating overview of analysis conditions.	60
Table 3.6	Results from Classifying Emotion using 20 fold CV on pressure-	
	location features for Touch (T), Touch + Gaze (T+G), and Touch	
	+ Gaze + Biometrics (T+G+B)	63
Table 3.7	Results from Classifying Emotion using 20 fold CV on Touch	
	Frequency (TF); Touch Frequency+Touch pressure-location (TF+T)	;
	Touch Frequency+Touch pressure-location+Gaze Frequency (TF+T	+GF)
	feature sets.	66
Table 3.8	Results from Classifying Participant using 20 fold CV on pressure-	
	location features for Touch (T), Touch+Gaze (T+G), Touch+Bio	
	(T+B), and Touch+Gaze+Biometrics (T+G+B)	68

Table 3.9	Results from Classifying Participant using 20 fold CV on Touch				
	Frequency (TF); Touch Frequency+Touch pressure-location (TF+T	');			
	Touch Frequency+Touch pressure-location+Gaze Frequency (TF+	Γ+GF)			
	feature sets.	69			

List of Figures

Figure 1.1	One complete iteration of the Human-Robot Interaction loop [110]:				
	(1) Human expresses emotion; (2) Robot recognizes Human				
	signal; (3) Robot expresses reaction to interpreted human ex-				
	pression; (4) Human recognizes Robot expression	2			
Figure 1.2	A collection of therapy robots designed to study the social				
	and physiological impacts of robots-as-companions on human				
	lives: (a) Paro; (b) Huggable; (c) Probo; (d) Haptic Creature;				
	(e) CuddleBot.	4			
Figure 2.1	(a) Top view of the CuddleBot skeleton. (b) Touch sensor,				
	pinned to foam substrate wrapped around the skeleton and cor-				
	responding to a No touch \times No motion \times No cover condi-				
	tion. (c) Full range of breathing motion used. (d) The fully-				
	covered robot; a covering of nearly identical material was used				
	in the study to facilitate quick condition changes. (e) The fab-				
	ric pressure sensor constructed out of EeonTex conductive fab-				
	ric $\langle www.eeonyx.com \rangle$, wired to an Arduino microprocessor.	16			
Figure 2.2	Mean gesture prediction accuracy rates with added pressure				
	noise when (a) varying substrate or cover in Study 1 and (b)				
	varying motion and cover on the same curved structure as in				
	Study 2. Each bar represents an average accuracy rate over 10				
	trials; error bars are omitted as Δ across trials $< 0.001\%$ in				
	each case	24			

Figure 2.3	A modified Hinton confusion matrix for gesture classification.	
	Horizontal (row) gestures are classified as the vertical (col-	
	umn) gesture. Saturation in non-diagonal squares represents	
	number of misclassifications	26
Figure 2.4	Cohen's d effect sizes of participant by gesture for each study.	28
Figure 2.5	Mean subject recognition rates by gesture and study over 10	
	trials; error bars omitted as Δ across trials $< 0.01\%$ in each	
	case. Study 2b refers to the 'in-motion, with cover' condition.	30
Figure 2.6	Top features as selected by Weka for each study. Classifica-	
	tion tasks are Gesture and Subject, by Location (C_x, C_y) and	
	Pressure features. Features selected at under 25% frequency in	
	20-fold cross validation are omitted	32
Figure 3.1	The experiment setup: participant sits comfortably supported	
	by pillows, facing the gaze tracker with her hand on the touch-	
	sensitive surface of a stationary robot. Biometric sensors are	
	worn around the waist (respiratory rate), thumb (blood-volume	
	pulse), and index and ring fingers (skin conductance) of the	
	resting hand. One camera captures eye movements and an-	
	other is raised on a tripod behind the participant to capture	
	hand motions over the robot. Both cameras have audio dis-	
	abled for privacy. When the participant pulls the rope, a ball	
	leading outside the room indicates to experimenters that the	
	emotion task is complete	48
Figure 3.2	Our robot was constructed from pliant plastic sheets actuated	
	by a pulley, covered with a custom-built sensor, and finally	
	wrapped in a furry fabric to invite touch. It remained station-	
	ary during the course of this study to eliminate any effect of	
	participant reaction to robot motion.	50

Figure 3.3	The data recorded during a single emotion task is referred to
	as a capture. To determine the effect of varying window size
	on accuracy rates, we tested at 2s, 1s, 0.5s, and 0.2s windows.
	Assuming that there may be similarities learned when trained
	and tested on directly adjacent windows, we also compared ac-
	curacy rates of data with and without imposed gaps to remove
	adjacent window instances.

- Figure 3.4 Classification performance by modality set, level of system knowledge (where *labels in* indicates with knowledge), window size, and with/without gaps. The Emotion row displays results from classifying emotion, with and without Participant labels. The Participant row displays results from classifying participant, with and without **Emotion** labels. For both cases, increasing the number of modalities improves accuracy rates, where the inclusion of biometrics provides the strongest classification rates. However, in the most rigorous test of emotion classification, Leave-one-out (LOO), all modalities perform at or near chance for all window sizes. Regardless of classification task, window size has a small or positive effect on accuracy rates, except for case of 2s-windows-with-gap. This is likely due to a high decimation of data, i.e. 2s-windows-withgap has 13% of the number of data instances as 1s-windowswith-gap.
- Figure 3.5 This is a bargraph of accuracy rates for every LOO test by participant. Most of the accuracy rates are roughly chance=25%, and low SD suggests the variance of this classification was low across window sizes. Notice a few interesting outliers: test sets including P22 and P05. P22 has much higher classification rates, suggesting that P22 may be similar enough to the group to have their emotion behaviours consistently identified. Conversely, P05's particularly low rate of classification suggests that P05 expresses emotions contrary to the group.

56

62

65

xii

Figure 3.6	Feature selection popularity of pressure-location features of	
	Touch and Frequency features of Touch and Gaze	67
Figure 3.7	Feature selection popularity by statistic. Pressure-based touch	
	features are most popular, followed by location-based touch	
	and gaze frequency features on location data also third, when	
	normalized against total number of features available by modal-	
	ity	71
Figure 3.8	Changes in individual's self-report of emotion after Neutral-	
	ization (start) and Emotion tasks (finish); N=14 for Stressed &	
	Relaxed and N=16 for Depressed & Excited. Overall, we see a	
	move from the origin to the representative quadrant. Stressed	
	and Excited show the strongest overall change along both Arousal	
	and Valence axes. Relaxed shows the least change with dis-	
	connected points referring to "no change" from neutral state	73
Figure 4.1	The rigid RibBit (Left) and fur-covered FlexiBit (Right) ex-	
	plore very different form factors using similar actuation princi-	
	ples and requirements. Both can be compressed without dam-	
	age, allowing for a more naturalistic haptic display	82
Figure 4.2	Experimental setup showing a participant touching the Flex-	
	iBit and rating behaviours. The screen's quadrants present the	
	four situation descriptions.	84
Figure 4.3	Close-up of the interface participants used to rate behaviours.	85
Figure 4.4	Waveforms of behaviours as designed by researchers	86
Figure 4.5	Mean behaviour ratings for FlexiBit grouped by the researcher-	
	designed behaviours (horizontal) and the situation for which	
	the behaviours were rated by participants (vertical). Researcher-	
	designed behaviours correspond with (a) to (h) in Fig. 4.4.	87
Figure 4.6	Pairwise comparison p-values (Wilcoxon) of behaviours (row)	
	for different situation conditions (col), sig. diff. are darker.	
	Notice RibBit(S-E, D): in the Depressed condition, <i>Stressed</i>	
	and <i>Excited</i> were rated significantly differently	88

Glossary

- **BRV** Breathing Rate Variability
- **BVP** Blood Volume Pulse
- DIY Do It Yourself
- **DOF** Degree-of-Freedom
- **FFT** Fast Fourier Transform
- **FSRS** Force Sensing Resistors
- HRI Human-Robot Interaction
- HRV Heart Rate Variability
- KS Kolmogorov-Smirnov
- LOO Leave One Out
- **RR** Respiratory Rate
- SC Skin Conductance

Acknowledgments

I am indebted to many people for their support.

Many thanks are owed to my supervisor, Dr. Karon MacLean, for believing in me and inviting me to join the SPIN lab, an experience which has facilitated so much learning and success not least of which are the relationships that have developed.

I am grateful to my second reader, Dr. Giuseppe Carenini, for the insightful comments and questions - this is a better work for it.

And to my labmates, collaborators, and friends: Paul Bucci, Jussi Rantala, Oliver Schneider, Hasti Seifi, Matthew Chun, Merel Jung, Soheil Kianzad, Dilan Uztek, Meghana Venkatswamy, David Marino, Alicia Woodside, Lucia Tseng, Ben Clark, Sazi Valair, Jeff Allen, Andrew Strang, and Michael Phan-Ba. I have much appreciated the valuable feedback, and thank you all for making the SPIN lab so fun!

Of course, I owe so much to my family: my mom, Hui Wang, and my sister, Alice Cang, for their unending love and support.

Finally, I am forever glad for my partner, Oliver Trujillo, for standing unwaveringly in my corner. Thank you for being my best friend.

This work was funded by the Natural Science and Engineering Research Council of Canada.

Dedication

To my mom, who has always demonstrated that love is so much more than words. Thank you.

Chapter 1

Introduction

Communication between people is made richer with affective clues. We pick up on each other's eye contact, vocal inflections, body language, touch behaviour as much as we do with words. Machine recognition of these kinds of cues improves the quality of Human-Robot Interaction (HRI) in collaborative tasks [36], intelligent tutoring systems [48], assisted driving [31], and the list goes on.

Using machines to recognize social *touch* specifically leverages natural human inclinations to express emotional closeness through physical contact and is gaining attention within affective computing [20, 33, 51, 93, 110]. As interpersonal touch encodes significant emotional content [43], investigating machine-sensed social touch is a stepping stone toward real-time emotion detection. One such application: the therapy robot pet takes advantage of the emotional communication in touch between human-animal without having to first address many of the complexities elicited in human-human interactions [78]. Therapy animals have long been shown to have physiologically measurable benefits on patients [10, 24]; for those who are unable to maintain long-term contact with pets—allergies, anxiety, cost—therapy robots have been employed with surprising success [46, 54, 103]. With better affective prediction, we can develop more naturally reactive therapeutic robots approaching that of a touch-based human-animal interaction loop as defined by Yohanan et al (2012) (see Figure 1.1).

Currently, there are obstacles to developing affective social touch recognition between a human and an animal-like robot pet, the biggest of which asks what pa-



Figure 1.1: One complete iteration of the Human-Robot Interaction loop [110]: (1) Human expresses emotion; (2) Robot recognizes Human signal; (3) Robot expresses reaction to interpreted human expression; (4) Human recognizes Robot expression

rameters of touch behaviour conveys emotional content and whether or not touch sensors can capture it. We focus here on those we think are significant and logical steps towards emotional communication in robot pet therapy. First, the robot's touch sensitive skin must flex with motion yet maintain data integrity under deformation. Second, classification of affective state in touch is less mature relative to some other modalities [4] and has fewer established techniques. Integrating touch with multiple channels known to contain affect increases our confidence in emotion detection. Third, even once affect classification is solved, our interaction is not complete without robot response. Researchers have studied emotions as exhibited in the behaviour of many different animals [12], however it's unclear if a robot simulation of affective expression is still human-identifiable as emotion. Meanwhile (fourth), we must be mindful of real-time viability throughout the entire interactive cycle – from sensing human touch, to predicting affective state, to developing appropriate robot responses – and consider ways to trim computational cost wherever possible.

In this thesis, we focus on the therapy robot pet application and describe a custom sensing mechanism as a method for extracting touch data, supported by three distinct studies. In the first, we compared classification of touch gestures under a variety of deformation conditions, allowing us to recommend a set of conditions that balances user preference with data quality. The second study collects emotive touch data, supported by gaze and biometric sensors. The multimodal approach helps us understand how users express emotion as well as compares affect classification accuracy between touch alone vs with support; this chapter concludes with design recommendations for an affective robot pet. To close the loop, we investigate emotional expressiveness in robot breathing where participants identify robot expression of affect while interacting with small single Degree-of-Freedom (DOF) robots performing a variety of breathing patterns. Finally, we outline the outcomes and impacts from this body of work and ground our findings in future work for furthering the therapeutic robot pet.

1.1 Background

Here we describe the bigger-picture background that leads to us to this work. More focused literature reviews are included in each forthcoming chapter.

1.1.1 Robot platforms

Multiple robot form factors attempt to imitate animal therapy success though it is not yet clear which characteristics generate the measurable physiological improvements in cardiopulmonary pressures, neurohormone levels, and anxiety in patients [24]. Some are plush, cuddly versions of larger mammals like Paro, a seal, or the Huggable, a teddy bear; others have no Earth-born analogue, like the furred, green Probo with a long articulated nose or the Haptic Creature, a round, furry object reminiscent of a cat/rabbit hybrid. Each robot has distinct sensory and actuation capabilities (see Table 1.1), but all are designed to allow for common pet



Figure 1.2: A collection of therapy robots designed to study the social and physiological impacts of robots-as-companions on human lives: (a) Paro; (b) Huggable; (c) Probo; (d) Haptic Creature; (e) CuddleBot.

interactions, including stroking and hugging.

Robot	Modalities	Actuation (DOF)	Special Features	Form	Surface
	sight (light v dark)	neck (2)	weighs 2.7kg		
Paro [46, 90]	sound (direction and speech)	front paddle (1)	pacifier	baby harp seal	soft white fur
	balance	rear paddle (1)	reacts to stroke and hit		
	touch \sim 2inch taxels	eyes (2)	pacemaker-friendly		
11 11 504 051	touch: temperature, force, electric field	neck (3)	recognizes 9 gestures		
Huggable [94, 95]	sight: video camera in eyes	eyebrows (2)	wireless connectivity	teddy bear	soft butterscotch fur
	sound: microphone	shoulders (2)	speaker for audio out- put		
	balance: Inertial mea- surement unit	ears (1)			
	sight: digital camera	eyebrows (4)	wifi-enabled touch screen		
	sound: microphone	trunk (3)	classifies 3 touches		
Probo [86]	touch: position sensors, temperature	mouth (3)	facial recognition	elephant-caricature	soft green felt
		head (3)	speech analysis		
		eyes(3)			
		eyelids (2)			
		ears (2)			
Haptic Creature [108]	touch: 60 FSRs over entire body	ears (2)	fiberglass shell	cat/rabbit-like form	soft brown synthetic fur
	balance: internal ac- celerometer	breathing (1)			
		purr (1)			
	touch: 256 taxel fabric	head (2)	WIFI-enabled		
CuddleBot [3, 19]	balance: accelerometer and gyroscope	ribs (1)	3D printed skeleton guinea pig-like form soft minky		soft minky
		spine (1)]		
		purr (1)			

Table 1.1: An overview of robot platforms intended for therapeutic use.

S

Paro The choice of emulating a baby harp seal exploits cuteness and recognizability, without being too familiar and pre-empting users expectations of behaviour [104]. Paro is fully autonomous with tactile sensing systems that allow it to recognize whether it is being *stroked* or *hit*, adjusting behaviour as appropriate [46]. The custom ubiquitous tactile sensors employed on Paro [90] are pictured as large palmsized pressure sensors and placed where touch contact is most likely – the head and back, while avoiding more difficult to manage locations like the joints. Use of Paro as a companion in care homes for senior citizens has been well-documented, particularly as a comfort animal surrogate for dementia patients [21, 35].

Huggable The Huggable is primarily designed for pediatric care and is intended as an augmented comfort object for children experiencing the stress of hospitalization [96]. An accompanying web-based logging program can be used to monitor the patient's distress levels through video and audio channels, as well as touch or behavioural data. The web interface also enables an operator, such as a therapist or caregiver, to control Huggable's many actuators and react to a patient's behaviour.

Probo Probo, also intended for engagement with children, relies mainly on facial expression of emotion. The cartoon elephant sports an articulated trunk as well as an interactive touch screen installed in the belly [86]. Like the Huggable, Probo requires an operator to communicate with the child where interaction is intended to evolve into a friendship [38].

Haptic Creature The Haptic Creature was developed to highlight touch-based interactions in social robot therapy, reducing visual cues as much as possible [108]. While there are multiple actuation avenues, simple regular breathing motions from the lifting fiberglass body plates is enough to elicit significant calming effects as demonstrated by biometric measures of reduced heart and breathing rates [88].

All of these robot platforms target therapy and are equipped with impressive kinematic abilities; our goals for complex and autonomous affective touch communication, however, require more sophisticated sensing and processing of affective touch signals. In fact, only Paro is actually intended and equipped to react to users in the absence of a human interpreter and operator. Our desired interaction, however, assumes a larger and more complex emotion set necessitating comprehensive full-body touch sensing as well as an embedded touch prediction engine.

1.1.2 The CuddleBot and CuddleBits

To build upon the developments of the Haptic Creature, other members within the lab created the CuddleBot, a new platform for study in therapy robot [3]. Improvements include increasing the touch-sensitivity to cover the entire body, more realistic motion with a 3-D printed, articulated full skeletal structure on 5-DOF, Wi-Fi connectivity for use with a web interface to define and adjust behaviours. The core structure is driven by modular, centrally positioned, rod-driven actuators as robot designers were mindful of keeping the kinematics easy to modify, enabling quick exploration into design alternatives. The CuddleBot, like Paro, is designed to be fully autonomous, but with a more complex on-board real-time affect interpretation engine able to interpret emotional expression in touch and reacting accordingly.

Many engineering parameters were considered when designing the CuddleBot, including designing for full modularity so that each degree of freedom could be developed and studied independently. To explore expressivity of individual actuation methods, a family of single DOF robots, dubbed the CuddleBits, was created [18]. Internal tools were also developed, enabling the sketching and fine-tuning behaviours to investigate the perception of emotionally expression of these simple machines [15].

For either robot platform, the affective prediction model, its mapping to a set of robot responses, and the human perception thereof remained undeveloped.

1.1.3 Sensing technology and classification techniques

Classification procedures depend on polling rate, resolution, format, and expected purpose of the data collected. If we only intend to distinguish between two touch gestures, say *stroke* and *hit* as in the case of Paro [46], the larger ubiquitous tactile sensor [90] in conjunction with a prediction system that detects pressure threshold is sufficient for use. For classification of more nuanced touch, however, not only are more taxels of higher resolution required, but also more sophisticated processing.

Sensing

High resolution sensors, such as those employed in robot grippers, are welldeveloped for very precise, dextrous manipulation and are successful for telemanipulation or even robot-assisted telesurgery [63]. However, this level of precision far exceeds our needs for affective or social touch gesture interpretation, and would be excessively expensive, both financially and computationally.

A survey of other touch sensors including Force Sensing Resistors (FSRS), gridbased pressure sensors (available at www.plugandwear.com), and fabric stretch sensors (www.vista-medical.com) revealed that they either must be affixed to a rigid substrate, were not capable of multi-dimensional stretch, or did not reflect multi-contact touch. Other requirements include fully-stretchable, full-body continuous taxel coverage, while being financially and computationally cheap. Unfortunately, existing sensors could not satisfy all criteria (requirements and touch sensors described in more detail in Section 2.1.3 and Section 2.2.2 respectively).

Classification

Classification procedures that make use of both pressure and location parameters have performed at up to 94% accuracy (chance 11%) with random forest models built on gestural touch data [33].

In contrast, machine recognition of affective touch alone does not perform nearly so well, with results at up to 48% accuracy (chance 11%) [4]. By enhancing unimodal touch data with information from other comparably nonverbal sources, however, we may be able to do much better.

Affective state directly influences physiological measures [32, 62] and has been classified with accuracies as high as 95% (chance 25%). Tracking gaze behaviour has also been promising in determining emotional state and attention or interest, leading to improvements for intelligent tutoring systems [48] and educational games [25], to name a few.

Affect recognition in touch may benefit from integrating multimodal support of gaze and biometric channels, both of which are associated with commercially available sensors and feature extraction software. As far as we know, there does not currently exist a study that triangulates even two of these three signals.

1.1.4 Interaction Styles

To complete the touch interaction loop, we require that on-board robot sensing, affect classification and motion response work together.

Looking at 1-DOF at a time, we begin by focusing on breathing behaviours due

to its range of emotional expressivity [13] as well as its proven ability to reduce symptoms of stress [88]. Section 4.1 delves into greater detail on robot behaviour complexity and believability.

We consider possible designs for the interaction model in two relevant interactions types, as defined by Sharp [89]: *Instructing* – where users issue instructions to a system – and *Conversing* – where users have a dialog with a system.

Instructing interactions can be conceived of as a continuously listening Robot waiting for Human direction. Once the recognition engine provides a prediction, the Robot performs the appropriate mapped behaviour and returns to a listening phase – completing one iteration of Yohanan's interaction loop (see Figure 1.1) where each iteration is completely independent. Producing appropriate responses requires that our prediction system perform with high accuracy and be able to gauge user intent with little room for error.

Conversing interactions are more suited for affective response: the Robot is continuously listening for Human input whereupon touch input triggers a prediction, generating a Robot behaviour. Upon detecting Human reaction to the earlier response, the Robot reflects on the previous behaviour and decides either to correct or continue the reaction.

While many human-human or human-animal interactions are well-adapted to instructing interactions, emotional communication is much better described in *conversing* style. Whether engaging with other humans or with animals, we assume some rate of error and are continuously evaluating and correcting behaviour based on increasing information. Although this more realistic model is outside the scope of this thesis, we lay the groundwork for modelling this kind of emotional human-robot conversation.

In either interaction style, interesting and believable human-robot interactions necessitate a large and complex behaviour set where each behaviour is human-recognizable as an emotional response. Unfortunately, generating emotional robot behaviours is difficult without more insight into the perception of platform-specific robot behaviours.

1.2 Approach

The overarching goal of this thesis is to explore and improve affect recognition in touch behaviour and expand its use for a variety of emotionally intelligent applications. We focus on social robot therapy because this application is easy to motivate for participants and naturally elicits human emotional expression.

To satisfy the sensing needs of a fully touch-sensitive robot pet in motion, we customized a Do It Yourself (DIY) fabric touch sensor of 1-inch taxels in a 10×10 array to be consistent with human social touch behaviours in terms of pressure range and polling rate [92]. The sensor reports pressure and location dimensions as required for gesture recognition [33] and is used in both gestural and affective touch studies (sensor construction is fully detailed in Section 2.3.1).

The literature reflects an assumption that there is emotional encoding within gestures [4, 33] and recognition engines of other robot platforms build on this scaffold. For example, Paro responds to *stroke* positively and *hit* negatively [46], and the Huggable and Probo respectively detect nine [96] and three [86] touch gestures to react to emotionally. While we acknowledge this is a reasonable approach, we choose to explore an independent relationship between gestural and emotion behaviour, electing instead to perform distinct collection procedures and report classification results separately.

Machine recognition of affect in touch remains largely unsolved: we have not yet reliably pinpointed the aspects or artifacts of touch that decode emotional state. However, since emotions are experienced multisensorially, we can borrow from the classification techniques of more mature sensing channels and introduce a multimodal approach by integrating gaze and biometrics support with touch. We then compare touch-only performance with touch plus additional modalities and discuss the use of multimodal sensing across a number of applications.

Finally, we evaluate the expressivity of breathing by designing a number of behaviours on the 1-DOF CuddleBit robots and compiling user reactions to them. By evaluating each stage of the interaction loop independently, we isolate requirements and limitations, gaining a deeper understanding of how to engage in the full HRI touch interaction loop. While out of scope of this thesis, these results begin to extend naive loop iterations to smarter behaviours with error correction and

improve the HRI experience.

1.3 Thesis Organization

The remainder of the chapters of this thesis are organized by the studies conducted.

In Chapter 2 (Gesture Classification), we describe the sensor used in both touch interaction studies and perform gesture classification. A major goal of this study was to assess impact on recognition performance of sensor motion, substrate and coverings. We collected six gestures most relevant in a haptic social robot context plus a no-touch control ($N_1 = 10$, $N_2 = 16$) and ran classification on a random forest model of 100 trees using Weka, an open-source machine learning program. Results allowed us to conclude that under realistic conditions (CuddleBot in motion, with foam substrate, and under a fur cover), recognition using our custom sensor is sufficient for many applications of social touch including affective or functional communication, from physically interactive robots to any touch-sensitive object.

Chapter 3 (Affect Detection) describes our multimodal collection and classification of emotional expression in touch (measuring force magnitude or pressure and location), gaze (location), and biometric (skin conductance, blood volume pulse, respiration) data. We asked participants (N=30) to relive intense emotional memories while touching a stationary, furry robot, eliciting authentically experienced emotions under controlled lab conditions. Each participant interacted with the robot while experiencing two opposing emotions: Stressed-Relaxed or Depressed-Excited. To better understand the data, we partition by level of system knowledge of participant information [labelled, unlabelled, leave-out] and across time-varying windows [2s, 1s, 0.5s, 0.2s], we found classification accuracy rates improve with increasing system knowledge of participant. Adding modalities of gaze and biometric data improve accuracy only when a participant's instances are present in both training and test sets. To address the computational efficiency required of dynamic and adaptive real-time interactions, we analyze relative subsets of our multimodal feature set, and provide design recommendations for our therapeutic robot.

The last stage of a single loop iteration requires robot reaction, investigated in Chapter 4 (Behaviour Sketching). We describe our design of breathing behaviours via direct waveform modification, displayed on the CuddleBits, one flexible, furry *FlexiBit* and the other a rigid, wooden *RibBit*. Emotive behaviours were designed by experimenters and interpreted by N=20 participants based on an arousal/valence emotion grid, specifically *Excited*, *Stressed*, *Relaxed*, and *Depressed*. Our findings indicate that these simple robots haptically conveyed emotion with success similar to that of more complex systems.

Finally, Chapter 5 (Conclusion) highlights the outcomes and impacts of each study and describes future work to improve the behaviour design in a human-robot conversation.

Chapter 2

Gesture Classification

Earlier robot constructions such as the Haptic Creature have used FSRS as the touch sensing system [109] requiring a firm, plexiglass construction that was heavy and offered insufficient degrees of freedom for convincing, complex motion. A full redesign yielded the CuddleBot [3] whose light 3D-printed skeleton necessitated an even lighter-weight sensing system. We needed a touch-sensitive skin that could flex around multi-DOF motion without losing data integrity, kept a low-computation profile to facilitate real-time processing, *and* was inviting to touch.

No sensor quite fit all of our requirements until we found inspiration in maker culture. Following existing DIY-sensor guidelines [77], we did a custom modification using commercially available conductive fabrics ¹in order to capture both pressure and location dimensions required for gesture recognition [33]. While this sensor is not presented as a contribution, we highlight its development in this Chapter. The same sensor is used to collect touch data in Chapter 3.

This chapter features a conference paper previously published at the ACM International Conference for Multimodal Interaction (ICMI'15) [19], presented here in full. Prior to attempting the more complex affect recognition, it was necessary to verify that the use of our custom fabric sensor in gesture recognition tasks demonstrated results consistent with the literature. This paper also highlights the noise concerns when engaging the CuddleBot robot in real-use. We present results from data collected under increasingly difficult conditions (stationary to moving; firm,

¹Sensor fabric purchased from (www.eeonyx.com)

flat substrate to soft, curved foam; no cover to high density fur) and recommend a configuration that balances performance with user preferences.

Abstract

Social touch is an essential non-verbal channel whose great interactive potential can be realized by the ability to recognize gestures performed on inviting surfaces. To assess impact on recognition performance of sensor motion, substrate and coverings, we collected gesture data from a low-cost multitouch fabric pressure-location sensor while varying these factors. For six gestures most relevant in a haptic social robot context plus a no-touch control, we conducted two studies, with the sensor (1) stationary, varying *substrate* and *cover* (n=10); and (2) attached to a robot under a fur covering, *flexing* or *stationary* (n=16).

For a stationary sensor, a random forest model achieved 90.0% recognition accuracy (chance 14.2%) when trained on all data, but as high as 94.6% (mean 89.1%) when trained on the same individual. A curved, flexing surface achieved 79.4% overall but averaged 85.7% when trained and tested on the same individual. These results suggest that under realistic conditions, recognition with this type of flexible sensor is sufficient for many applications of interactive social touch. We further found evidence that users exhibit an idiosyncratic 'touch signature', with potential to identify the toucher. Both findings enable varied contexts of affective or functional touch communication, from physically interactive robots to any touch-sensitive object.

2.1 Introduction

Words can sometimes be inefficient for communicating instructions or affective content. In many contexts, touch may be the best modality for conveying directive and emotion: imagine how informing someone to get out of the way quickly and clearly with one simple touch. To harness this communication channel, social robots working in tandem with humans must recognize the same haptic language that we use, of which gestures and affect are key components.

We focus on exploring the range of touch gestures detectable by a custom-built flexible fabric pressure sensor and evaluating the added noise from curvature, motion and material cover. Using common machine learning techniques, we highlight salient features of touch for recognizing both the type of gesture being performed, and the person performing the gesture—both the 'touch' and the 'toucher'.

Reliable gesture recognition is an important step towards further research in the field of affective touch. A strong foundation of research on gesture may allow us to detect the toucher's emotional state [50]. Until recently, this kind of research was difficult, as touch sensors were not easily deformable nor cheap in both price and computational resources. Our 10×10 sensor has 100 fingerpad-scale taxels recording pressure and 2-D location data, and we use a random forest classification method to approximate in situ recognition rates.

We first collected touch data for a set of six validated touch gestures [110] plus one control on a stationary sensor under a variety of substrate stiffnesses and coverings. We then mounted the same sensor on an actuated robot skeleton and collected similar data while varying the sensor's covering and motion (Fig. 2.1). Recognition rates were within 80–95% for all conditions we tested (chance 14.2%), a level of accuracy which will suffice for many purposes and is enough to merit empirical comparison to human recognition ability in future work. At the same time, we found individuals' touch signatures were idiosyncratic enough to permit identification of toucher within this sample, at an accuracy rate similar to that of the gestures themselves.

2.1.1 Questions and Contributions

We wished to learn:

- **Q1:** How accurate is our flexible fabric sensor in predicting gesture and differentiating between users?;
- **Q2:** How does sensor performance hold up under deformation due to curvature and motion, such as that produced by a zoomorphic social robot?; and
- **Q3:** Is real-time gesture recognition computationally viable?

With 20-fold cross validation on random forest models, we contribute initial results of:



- Figure 2.1: (a) Top view of the CuddleBot skeleton. (b) Touch sensor, pinned to foam substrate wrapped around the skeleton and corresponding to a *No touch × No motion × No cover* condition. (c) Full range of breathing motion used. (d) The fully-covered robot; a covering of nearly identical material was used in the study to facilitate quick condition changes. (e) The fabric pressure sensor constructed out of EeonTex conductive fabric (www.eeonyx.com), wired to an Arduino microprocessor.
- deployable accuracy in gesture recognition (6 gestures + control): 91.4% on a firm, flat surface, 90.3% on a foam, curved surface, and 88.4% on a foam, curved, moving surface;
- differentiating toucher at 88.8% accuracy (n=26);
- factors underlying recognition performance;
- feasibility of real-time gesture recognition.

We also make our data and analysis publicly available².

Our study compares gesture recognition performance across a variety of conditions that approach real-time dynamic gesture recognition. Toucher recognition accuracy shows promise for incorporating personalized responses to an individual touch signature.

2.1.2 Applications

Accurate gesture recognition on a fabric touch sensor opens up gesture-based controls on any electronic device. For example, patients with limited speech could use a smart blanket with gesture recognition capabilities for comfort or health-

²All collected data and select analysis can be found at (www.cs.ubc.ca/labs/spin/data)

reporting purposes. In the context of social robots, a sensor that can wrap around any irregular form could be used as a touch-sensitive skin. Outside of explicit gesture recognition, pressure-sensitive hospital sheets could alert caregivers of bedsore risk.

In a behavioural education context, a soft touch-sensing playmate capable of recognizing touch signatures may use this data to interpret and influence emotional state [50]. Such a robot could aid students testing on the autism spectrum by responding to anxious or agitated strokes with slow, soothing, regulated breathing—a behaviour shown to have calming benefits [88].

2.1.3 **Detailed Requirements**

Our sensing requirements are dictated by a zoomorphic robot, affectionately dubbed the CuddleBot, that invites touch with a soft furry body. Since a user will expect to interact with the CuddleBot via touch, having a full-body sensor that deforms with robot motion is required.

Movement and elasticity: The sensor must be highly flexible, somewhat elastic, and perform well while mounted on non-rigid and/or actuated surfaces.

Pressure range: Based on a preliminary survey of these touch pressures, we determined that our sensor needed to register touches between 0.005 and 1 kg. This range is appropriate for light tickles to heavy pats.

Multitouch: Multitouch capability allows us to compute varying pressure over an area, differentiating touches like *constant* and *pat* from *tickle* and *scratch*.

Resolution and computational cost: Taxel resolution, sampling rate, and computational cost must be balanced to achieve usable recognition accuracy. For realtime, our computational cost is dominated by sensor polling and grows with the number of taxels per grid edge. Our recognition tasks and feature selection explicitly analyze the differences between frames. In this case, accuracy plateaus with fingerpad-scale taxels, when sampled fast enough to capture voluntary movement (peaking at 10Hz [92]). We must be able to recognize changes in pressure and localized hand motions up to this frequency.

Single-fingerpad resolution (≈ 2 taxels per inch) could capture small fluctuations; however, our gestures (not including our control *no touch*) either involve the flat or palm of hand (*constant*, *pat*, *rub*, *stroke*), or tend towards quickly crossing many taxels (*tickle*, *scratch*). This suggests that using statistical features that emphasized the changes from frame to frame could be used to achieve reasonable classification rates even at ≈ 1 inch taxels [33].

2.2 Related Work

We situate our work in the context of social robotics and affect-encoding social touch. Gestural touch has been identified as a key component of human-robot cooperation [7]. However, the semantics of that touch is conveyed through nuance. For example, the same gesture could halt, contribute or modify another person's behaviour [7] depending on the emotional content inferred from pressure dynamics [50].

2.2.1 Social & Affective Touch Communication

In collaboration with human workers, robots employed in a laboratory or workshop setting presupposes a lexicon of social touch for operational interactions [36]. To ensure safe and effective communication, Gleeson et al identify the requirements of both a comprehensive gestural dictionary and lightweight sensing technology. The intimate nature of collaborative robotic household help emphasizes the importance of affect detection for social robots in this context [2, 79].

Previous work revealed correlations between gestural social touch and emotional communication [44, 50]. Humans recognize the affect encoded in gestural touch [43, 44], suggesting that machine recognition of emotional state can be achieved with sufficient sensing technology and clever feature extraction.

Much of the current work on social touch recognition uses a sensor worn on a static human or robotic arm [50–52, 93]. The collected data and signal processing procedures may not account for the added deformation noise of a soft-tissue zoomorphic robotic form in motion.

The use of animals [24] and interactive robots in animal form (such as Sony's pet-dog AIBO [9, 97], the seal-shaped PARO [46, 64, 91, 104]) suggest potential benefits in therapeutic use. Other touchable social robots include the teddy bear-like Huggable [94]; and Probo [85], which does not have a recognizable animal

analogue. However, while real pets respond to complex touch commands anywhere on the body, this has been difficult to achieve without a generalized touch-sensitive skin.

In trying to establish zoomorphic robots as an emotional agent [33, 110], touch sensing strategies have included fur-level conductive threads, extensive biometric data, gyroscopes and accelerometers, to name a few. While this cavalcade of sensing produces encouraging results for social gesture classification [33, 110], it is far from the light-weight system required for automatic, real-time recognition.

An unexpected result emerging from social touch recognition is the demonstrably higher accuracy results for within-subject classification over between-subject [33, 51]. Leveraging this result may allow us to use touch behaviours to identify individuals and thus, recognize the nuances of an individual's "touch signature" to better predict touch gestures and, eventually, basic emotional content.

2.2.2 Flexible Pressure-Location Sensors

Real-time classification of social touch gestures on a flexing, noisy surface requires that we have manageable signal processing while retaining the ability to represent pressure and location.

Here we examine the suitability of existing sensing technology and recognize their influence on our custom build. We do not present our sensor as a contribution.

While many highly accurate pressure-location sensors exist, such as those developed for robot grippers used in dexterous manipulation [82, 100], these tend to be insufficiently flexible, overkill in terms of resolution, and considerably too expensive for the objectives outlined here.

Other work has used Force-Sensing Resistors (FSRs) affixed to a hard shell [110]. This reduces the need to calibrate for sensor drift over continued use, however, the trade-off non-aesthetic tactility, and difficulty in detecting touches between sensors—limiting rendered motion [7, 20].

Stretch sensors designed for medical purposes by Vista Medical³ is the foremost inspiration for our custom sensor. However, Vista's sensors recognized only pressure without localization and did not have multitouch capability.

³Stretchable sensors can be purchased commercially from Vista Medical \langle www.vista-medical. com/subsite/stretch.php \rangle

Several multitouch, flexible fabric sensors are available [52]. However, flexibility alone does not afford a full range of motion; it must be able to stretch and deform to approximate animal skin.

The design and sensing capabilities described by Flagg et al [33] informed many of our requirements and suggested that the bulk of the recognition accuracy could be achieved by the "below surface" sensor alone. However, Flagg's study did not consider the full design space of a robot in motion including a non-sensing fur and a variety of configurations. To evaluate how much information is compromised under these conditions, we applied a variety of realistic use noise sources to the sensor, both directly and indirectly.

2.3 Studies

We hypothesized that:

- **H1:** *gesture recognition rates will decrease with noise-creating factors*—allowing us to rank these factors' impact on recognition performance, and their interactions therein.
- **H2:** *variability in gesture execution will be higher between subjects than within subjects*—giving rise to the potential of differentiating individuals based on personal touch signatures.

2.3.1 Apparatus

We constructed a sensor by layering two squares of conductive EeonTex⁴ Zebra fabric, aligned at 90 degrees, with a plastic standoff mesh separator and a sheet of EeonTex SLPA 20k Ω resistive fabric. Resistance value across a given taxel drops when pressure is applied, compressing the mesh separator so the conductive layers more closely approach each other. A circuit is constructed using an Arduino Mega microprocessor. Each fabric stripe is connected to a single I/O pin: the top layer is connected to analog input pins, and the bottom layer is connected to digital output pins (Fig. 2.1(e)).

⁴Sensor fabric purchased from (www.eeonyx.com)
The sensor is polled by sequentially sending a voltage through the bottom layer's digital pins. The analog pins read current; resistance (and hence current) varies with pressure.

Preliminary testing of our sensor using stationary weights showed that under ideal conditions, we were able to achieve a touch weight range of 0.005–1kg using 1k Ω resistors. Under the most severe conditions, lighter touches were lost in the dense fur; at the heavier end, touches were equalized by the yielding foam substrate. For Study 1, the curved-foam substrate with thick fur cover was the most obscuring condition; for Study 2, this was the cover condition with bot in motion.

Dynamic range is modulated through choice of resistor value. We found that values greater than $1k\Omega$ allowed our sensor to register greater forces, but lost resolution; conversely, lower values gave greater granularity in recognizing very fine touches, but were too vulnerable to saturation at commonly applied force levels. The same sensor and microprocessor set were used in all studies described here.

2.3.2 Methods

Our two studies assessed how realistic conditions impacted sensor data and hence recognition accuracy; gestures and data collection procedures were unchanged.

Gestures and Sampling

We selected gestures from Yohanan et al's touch dictionary [110], choosing items most appropriate for human-animal interactions [33]. The sensor was placed on a table in front of a seated participant, a reference sheet with very general definitions for six selected gestures and one control was provided (Table 2.1). Participants were instructed to interpret each gesture as they saw fit; no further performance clarifications were provided.

A frame consisted of pressure data from all 100 taxels in the 10×10 grid. We collected 10 seconds of continuous hand touch data at 54 frames per second for each combination of gesture and condition, randomizing gestures and conditions wherever possible.

Gesture	Suggested Definition
no touch	no contact with the sensor (control)
constant	touch contact without movement
pat	quick & gentle touches with the flat of the hand
rub	moving the hand to and fro with firm pressure
scratch	rubbing with the fingertips
stroke	moving hand repeatedly
tickle	touching with light finger movements

Table 2.1: Touch gesture instructions as provided to participants.

Study 1: Cover and Substrate on Static Robot

We first measured gesture recognition for the static (unmoving) case, to assess impact of the sensor's substrate stiffness, curvature and covering thickness in absence of movement noise. This produced a factorial design of $4 \times 3 \times 7$ (*cover* × *substrate* × *gesture*), using gestures listed in Table 2.1.

Cover: The fabric's pile or density varied from no cover (participant touched sensor directly) to a very long, thick synthetic fur. Minky (a short furry fabric generally used for baby blankets), and a longer-furred fabric comprised intermediate variations.

Substrate: The material underneath the sensor consisted of a firm, flat surface (sensor affixed by velcro to a table); a spongy foam, flat surface; and a spongy foam, curved surface. In cases with foam, the sensor was pinned directly to the foam substrate.

To minimize sensor reading disturbances due to transitions (i.e., unwrapping and replacing the sensor on/off the robot body), we blocked our design on the $cover \times substrate$ conditions. Condition order was randomly generated for every participant, and gesture order was further randomized over each condition set. All participants completed all twelve masking conditions, with each generating 48 2s sample windows per gesture. A study session took approximately 50 minutes to complete. 10 volunteers (4 female, 6 male) were compensated \$10 for their time.

Study 2: Stationary vs Moving Robot

Our second study focused on the impact of the robot's breathing movement. We varied *cover* \times *motion* \times *gesture*, for a 2 \times 2 \times 7 factorial design. Factors consisted of *cover* = {cover, no cover}, *motion* = {breathing, not breathing}, and *gesture* = {set of seven gestures}. Each participant performed each condition combination twice in a randomly generated order.

In the breathing condition, the sensor was attached to the CuddleBot, a catsized robot designed for therapeutic use. Fig. 2.1(a-b) shows the naked skeleton and the sensor pinned to the foam intermediary. The robot's 'breathing' motion was created by extending and contracting the paired rib assemblies in a 14° arc from the spine at 0.5Hz (Fig. 2.1(c)). We draped and pinned fabric over the sensor, approximating a full fur jacket for condition randomization while limiting sensor disruption (Fig. 2.1(d)),

Each session began by asking the participant to interact freely with the covered, moving robot for 1 minute to reduce novelty. Each condition was then presented twice, in random order, for a total of $((2 \times 2 \times 7) + 1) = 57$ trials. 16 participants (10 female, 6 male) were compensated \$5 for the 30 minute session, each providing 32 2s samples of each gesture for every condition.

2.3.3 Analysis and Results

We discarded the first and last second of each 10s gesture capture and divided the remaining 8s into four 2s windows. The 2s window (at 54Hz) was chosen to allow each gesture some periodicity; all gestures fit completely within 1s (Flagg [33]). Given the challenge of determining gesture boundaries in a realistic, real-time setting when a motion is steadily repeated, a 2s window allows capture of *at least* 1 complete gesture cycle.

To account for translatory gestures, we also calculated a centroid (average geometric centre) weighted by the pressure reading for each frame. Centroids were defined by row C_x (Eq. 2.1) and column position C_y (Eq. 2.1 with *i* and *j* indices



Figure 2.2: Mean gesture prediction accuracy rates with added pressure noise when (a) varying *substrate* or *cover* in Study 1 and (b) varying *motion* and *cover* on the same curved structure as in Study 2. Each bar represents an average accuracy rate over 10 trials; error bars are omitted as Δ across trials < 0.001% in each case.

reversed):

$$C_x = \frac{\sum_{i=1}^{10} \sum_{j=1}^{10} i * pressure(i, j)}{\sum_{i=1}^{10} \sum_{j=1}^{10} pressure(i, j)}$$
(2.1)

We calculated weighted *pressure* by summing readings across each row, multiplying by index, and dividing by the unweighted frame sum (the sum of the full frame sensor reading). Repeated for each column, this provided a tuple of frame sum and centroid per frame.

As a "baseline" for both studies, we sampled sensor frames in the absence of gestures. In Study 1, each of the 12 (4 *cover* × 3 *substrate*) condition sets contributed 4320 frames; in Study 2, each of the 4 (2 *motion* × 2 *cover*) condition sets contributed 6912 frames. To establish the effect of noise under each condition, we ran MANOVA over three frame-level dependent variables: pressure, C_x -coordinate, and C_y -coordinate. In all cases except one⁵, all three variables showed significant differences at the p < 0.001 level. This indicates that the sensor is sensitive to

changes in these conditions.

The data fails the Shapiro-Wilks test of normality; however, visual inspections of residual Q-Q plots did not reveal any systematic patterns. Together with our large sample size (n > 4000 frames per condition), we proceeded with the normality assumption, alert to risk of inflated Type I error.

The six gestures (omitting no-touch data) were then compared with each other under the conditions of each study. MANOVA over the same three metrics (pressure, C_x , C_y) showed that gesture and participant combinations were statistically significant (p < 0.001). Differences in participant touch were detectable at frame level.

We calculated seven features across these three dimensions (frame value, C_x , C_y) for each 2s window for a total of 21 features. For each dimension, features are {maximum, minimum, mean, median, variance across all frames, total variance within the 2s window, area under the curve}. Condition variables (*curvature*, *fur*) or (*cover*, *motion*) make up the other features. *Participant* labels were included for *gesture* predictions and vice versa.

Each capture produced four 2s windows, providing repetition for training. Pairwise comparisons of all within- and between- capture windows generated two binomial distributions for statistically significant pairs using two-sample Kolmogorov-Smirnov (KS) tests. Permutation testing [37] using the KS test statistic did not detect a statistically significant difference (p = 0.214) between the distributions. This is consistent with our observations of participants varying touch behaviour both between and within captures.

We used Weka, an open-source machine learning application to classify gestures [41]. Flagg's comparison of random forest and a number of other algorithms showed that random forest performed best in gesture recognition of this kind [33]. We ran k-fold Cross Validation (CV) on Study 1 participant data for $k = \{5, 10, 20, 100\}$ and found less than 1% improvement between 20- and 100- folds. While this CV technique does ensure that any one instance is included in the test or training set and not both, it cannot promise subject-independent classification. Running Leave One Out (LOO) classification yielded slightly improved results but we were

⁵For Study 2, the condition of *with-cover*×*with-motion* under *no touch* did not show statistically significant differences in C_x data.



Figure 2.3: A modified Hinton confusion matrix for gesture classification. Horizontal (row) gestures are classified as the vertical (column) gesture. Saturation in non-diagonal squares represents number of misclassifications.

cautious to the inflated bias [58]. All reported classification performance is therefore based on the slightly conservative 20-fold cross validation of random forest models. Accuracy is defined as the percentage of data instances that are correctly classified.

Gesture Classification by Condition

H1: Gesture recognition rates will decrease with increase in noise-creating factors—*accepted*.

Comparing classification under Study 1 conditions (static surface), we found highest recognition accuracy with no cover on the firm, flat substrate case. Lowest performers were dense fur and curved, foam substrate. In Study 2 (dynamic surface, heavy versus no cover), conditioning across each of surface and motion factors had minor effect recognition rates (all $\approx 88\%$).

With models trained on individual, Study 1 showed little change in gesture

prediction rate compared to all-data models. Study 2 individually-trained results are more similar to other studies, which also report training on single-condition data [33, 52, 68, 98].

Cover-substrate-motion: Fig. 2.2 shows overall gesture recognition accuracy by study and condition set.

We assessed relative noise levels by calculating effect sizes of significant conditions. Cohen's d reveals a large effect ($|d| \ge 0.8$ [23]) with the introduction of curvature (vs *no substrate*) and fur and short minky (vs *no cover*) in Study 1. Large effects ($|d| \ge 0.8$ [23]) from Study 2 were from introducing the cover (regardless of motion), and from the combination of having motion and cover. Interestingly enough, adding motion by itself produced a very low effect ($d \le 0.08$). Further investigation into the interaction between cover and motion on pressure readings included Tukey's HSD of adjusted p-values to clarify the significance of stratified factors. While all other combinations remained significant at p < 0.05, the case of varying motion in the presence of a cover was alone insignificant at p.adj = 0.7.

A confusion matrix (Fig. 2.3) indicates how gestures were misclassified. In both studies, the most-misclassified was *tickle*.

Participant: We classified gestures with models trained by participant. In Study 1, mean accuracy was 89.1% $(\max=94.6\%)^6$. Models trained on all Study 1 data were accurate at 90.0%, i.e. within 1% of the mean accuracy of the individual-trained models. This indicated that training on participants did not improve recognition when data was not conditioned on noise-creating factors.

For Study 2 (fewer noise factors) we found a greater effect for models trained on participants (mean=86.5%, max=97.3%)⁷. Training across all data gave 82.1% accuracy.

The *motion* \times *cover* condition had an overall 79.4% recognition rate. Training on the subset of data with the most challenging conditions (in-motion, with cover) still produced a higher recognition rate when using individual-trained models (mean=85.7%, min=73.7%, max=95.1%).

⁶Study 1 gesture recognition accuracy by participant: P1-93.0%, P2-83.8%, P3-85.0%, P4-92.6%, P5-93.2%, P6-88.0%, P7-94.6%, P8-91.7%, P9-86.0%, P10-83.4%

⁷Study 2 gesture recognition accuracy by participant: P1-90.2%, P2-86.6%, P3-86.6%, P4-91.1%, P5-81.3%, P6-86.1%, P7-84.8%, P8-79.5%, P9-95.5%, P10-79.5%, P11-90.2%, P12-93.8%, P13-97.3%, P14-83.0%, P15-79.5%, P16-79.5%



Figure 2.4: Cohen's d effect sizes of participant by gesture for each study.

We compared mean pressure of gesture behaviours by individual versus that of the entire pool (i.e. how P1 performed *scratch* versus how all participants performed *scratch*). All incidences were significant at p < 0.05 (Cohen's d effect sizes reported in Fig. 2.4).

Toucher Recognition

H2: Variability in gesture execution will be higher between subjects than within subjects—*partially accepted*, for the case of data compared within the same noise conditions.

The ability to recognize toucher may have great impact on reading emotional state. We compared performance in participant classification for models trained across the entire dataset, with those trained on the 6 meaningful gestures of our *gesture* set (omitting *no touch*). We also look at accuracy rates on data collected in the most realistic condition (in-motion, with-cover).

Recognition rate by study: We compare recognition rate by study and gesture in Fig. 2.5. Study 1 achieves an overall accuracy rate of 78.5% (chance 10%), but for models trained by gesture, a mean of 87.9%. The highest contributing gesture is *constant* at 92.7%, followed by *pat* at 88.9%.

Using all Study 2 data, participant recognition was 80.3%. Training by gesture again showed recognition improvement; *constant* was best at 93.8%, followed by *pat* at 89.8% (mean, all 6 gestures: 85.4%).

Conditioning on only the *in-motion, with cover* factor, referred to in Fig. 2.5 as Study 2b, average recognition rates of participants are 89.8%. Further splitting data to additionally train models by gesture does not provide additional improvement in mean performance (85.2%); this time *pat* is the highest performer (93.8%) and *constant* a close second (90.6%).

We again refer to effect sizes (Fig. 2.4) to consider the role of pressure in participant recognition; individuals making different gestures exhibit considerable variation in pressure patterns.

2.4 Discussion

We discuss our findings in direct response to questions posed in Section 2.1.1.

Q1a: Potential accuracy of sensor in gesture recognition

Unsurprisingly, we found the highest recognition rate (94.8%) for the case of no covering and a flat, stiff, stationary surface (Study 1); these are the least demanding conditions and the ones we expected to perform the best.



Figure 2.5: Mean subject recognition rates by gesture and study over 10 trials; error bars omitted as Δ across trials < 0.01% in each case. Study 2b refers to the 'in-motion, with cover' condition.

In evaluating the degree to which noise factors degraded performance, we expected the noisiest conditions to be in Study 2: moving, curved, springy surface under a heavy fur cover. This achieved 88.6% recognition rate of our 6 gestures and 'no touch', among the lowest we observed. However, at just under 90%, this value is still usably high. Further work is required to assess the impact of nonuniform motion, as well as unknown gesture segmentation boundaries in lesser controlled conditions.

Q1b: Potential accuracy of sensor in user differentiation

Our studies show that the ability to pick a particular 'toucher' out of a known group varies by gesture. A priori knowledge of a condition also improves prediction accuracy, jumping from 80.3% trained over all data to 89.8% when trained on *in-motion, with-cover*, the noisiest condition. To see how this may change over the various gestures, we refer to Fig. 2.5 which ranks *constant* and *pat* as most identifiable. Fig. 2.4, which compares the effect size of pressure reading by participant and gesture, reveals that there are many large effects for *constant* gesture. This focus on pressure suggests that there may be revealing variations in individual 'heaviness of hand'.

Q2a: Impact on accuracy due to cover, substrate and motion

Over our two studies, we examined variations in cover thickness, substrate stiffness

and curvature, and motion. Summarized in Fig. 2.2, we now discuss the impact of these factors individually.

Cover: The effect of a cover on classification performance is significant; more so than the underlying motion (as noted by Section 2.3.3). Fig. 2.2 further illustrates this. Regardless of whether we partition our data by *cover* on/off or *motion* present/absent, we achieve gesture recognition of at least 88.1%, 6% higher than training overall (82.1%).

The pressure applied over a denser, heavier fur cover may muffle some of the lighter touches and degrade transmission of touch pressure and/or location, thus confusing some gestures.

Another possible explanation could be from added familiarity that the cover affords. For example, according to one subject, "When it had the fur on, I had a more pleasant experience...Without the fur, I found it difficult to touch it." (S7) This opinion was expressed in some form by 10 of 16 Study 2 participants. More research is needed to determine if the fur invited more naturalistic touching.

Substrate: Compared to a flat, hard surface, a flat foam substrate decreased recognition accuracy by about 1% (Fig. 2.2a). It had slightly less impact than curvature or, comparing to Study 2, than motion. Given the sensor's piezoresistive construction, we anticipated the effect of firmly compliant backing to be small; this finding confirms that a somewhat springy underlying surface (helpful for conveying the sense of an animal body as well as a pleasant tactility) is feasible under a largebody touch sensor.

Motion: The relatively small effect size of motion in raw frame data is unexpected. However, in the context of Tukey's HSD results (with a cover, the motion effect is insignificant), we gain some further insight into just how small the effect of regular periodic motion is, and we confidently rank motion noise behind that of a cover.

This is very promising for the larger premise of reliable touch sensing on a flexing surface.

Interaction of motion and cover: There is a large effect size for the interaction between *cover* and *motion*, which is absent in recognition performance conditioned on added noise factors (Fig. 2.2). This consistent improvement over training on all data (overall at 82.1%) suggests that these large effect sizes of noise interference

			Mean		Median		Мах		Min	Total	variance	-	Variance
		Ρ	L	Ρ	L	Ρ	L	Ρ	L	Ρ	L	Ρ	L
Study 1	Gesture												
	Subject												
Study 2	Gesture												
	Subject												

Figure 2.6: Top features as selected by Weka for each study. Classification tasks are Gesture and Subject, by Location (C_x, C_y) and Pressure features. Features selected at under 25% frequency in 20-fold cross validation are omitted.

have little effect on recognition as long as we train and test on the same condition.

Q2b: Gesture Recognizability:

Gesture confusion patterns reveal a considerable range of misclassification (the more saturated cells in Fig. 2.3). In Study 1, the most commonly misclassified gesture is *rub* as *tickle*; in Study 2 *scratch* is most misclassified as *rub*. Both pairs are commonly executed as quick back-and-forth motions. This may be related to relative gesture pressure by individual: gestures like *constant*, generally more stationary, are predicted consistently and also indicate a larger effect size by pressure (Fig. 2.4). Quick motions being lost in the heavier covering may also contribute to these errors.

Q3a: Feature Utility

Making computational economy is the key to real-time recognition. Prioritized feature selection allows us to focus on high-performing dimensions. To help understand relative feature utility in our recognition tasks, we used Weka's Attribute Evaluator function to find the highest-weighted features for the random forest model (Fig. 2.6).

The feature set with the greatest ability to differentiate gestures related to pres-

sure variance; meanwhile, location variance facilitated *toucher* recognition. People's touch signatures may vary more in physical location range, but, a gesture may be better characterized using pressure when *toucher* is known (Fig. 2.4).

These results suggest that using a subset of the features described here could increase computational efficiency, depending on the priority of recognition task needed and the variance exhibited by an actual data pool. Meanwhile, evaluating the performance of a reduced feature set is difficult due to the lack of a benchmark for comparing accuracy rates [51].

Q3b: Computational viability of real-time gesture recognition

The conditions evaluated here approached realism in some respects, specifically that of sensor covering, substrate, and underlying motion. Our post-hoc analysis indicated that a modern microprocessor could keep up with both sampling and recognition.

Our setup fell short of realism in at least one important factor: people are unlikely to perform distinct, discrete gestures with well-defined boundaries. A different computational architecture will be required to handle this problem (a topic of ongoing work). However at present, computational load is dominated by sampling rather than recognition, an overhead cost that will not necessarily change with real-time use (unless more selective sampling can be employed based on observed patterns of touching). It is thus quite likely that a more capable recognition engine will also be feasible with comparable computational resources. In situ real-time recognition may be better approximated by speaker-independent Leave-One-Out (LOO) sliding window. Our work uses k-fold CV as the more conservative accuracy rate (as compared to LOO), as we do expect a calibration process in which speaker behaviour is learned. Until we optimize, the best window size is unknown.

2.5 Conclusions

The results described here represent an initial feasibility assessment of the impact of flexing surfaces on gesture recognition performance. We found recognition rates from 80–95% for optimal to noisy conditions when distinguishing between social touch gestures relevant to interacting with a small touch-centric robotic entity. We further found an ability to distinguish individual toucher at a rate of 78.5% and

80.3% in Study 1 and Study 2 respectively. In the noisiest case (also the most realistic), training by condition increased participant recognition accuracy to 89.8%. The next step is evaluating more comprehensive sets of movement conditions.

The implication of a sensing system able to detect both individuality touch and toucher is considerable. For example, a sensor able to differentiate between users could provide a personalized set of experiences or controls.

Further, identifying the touch brings us closer to differentiating affective intent [50]; identifying toucher may allow us to qualify their touch behaviour. A sensor loaded with a personal touch profile could determine how far an individual deviates from that profile on a given day, and infer emotional status. To build such a profile, it will be important to establish the dimensions of a touch signature.

2.6 Future Work

We foresee many ways in which to extend this work.

More extensive movement conditions: The present study employed steady periodic motion of an underlying surface for a flexible sensor. A more general, and potentially challenging, environment will include irregular and unexpected motions.

Continuous gestures: The single-gesture samples of this study removed the need to segment data in pre-processing. In future, an algorithm will not know of gesture boundaries or length a priori, and will need to handle the case of seamlessly transitioning gestures.

One approach is to run several sampling windows of different length to search for varying touch activations at the cost of increased computational load. Future work needs to explore this and other architecture to determine a strategy to optimize for computational efficiency.

Pragmatic gestures: In this study, participants were instructed to perform a particular named touch gesture, but not with communicative intent or emotion context. The semantics of a "natural" touch will be dependent on context of situation and the user's own state; to determine communicative intent, it may be necessary to observe other factors as well.

Our participants often varied in how they interpreted a given gesture, both between participants, and individually between and within conditions. For the latter, we suspect users may have performed more authentic gestures on the moving, furcovered robot than when it was flat, stationary and/or uncovered. We also observed differing touch behaviour from the beginning through the end of one capture, but our sensing mechanisms are unable to distinguish these cases.

Gesture stabilization and system interactivity: Finally, with more efficient algorithms deployable in realistic conditions, we plan a longitudinal study of long-term interactions in natural settings to investigate how individual gestures change over time as a toucher learns to interact with the sensing system.

2.7 Acknowledgments

We would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), SurfNet and the UBC SURE program for providing partial support of this work. We also thank Michael Phan-Ba for engineering the firmware and robot control, Philippe Kruchten for his mentorship, and finally, Oliver Schneider for the excellent title.

Chapter 3

Affect Detection

Automatic affect detection allow machines to extend beyond communication via instructional gestures to access more contextual cues. Emotions are expressed through many physical channels, including (but not limited to) physiology [62], facial expressions [60], eye behaviour [48, 74], speech [73], vocal prosody [59], and even touch [43]. Since machine recognition of emotional expression through touch is still largely unsolved and as yet unreliable [4], we use multimodal sensing to triangulate affective signals by leveraging gaze and biometric support of touch. In Chapter 2, we established that the custom-build sensor performed at literature-levels for gesture classification (80%–95%), so we proceed with the same sensor here.

In real-use case, a therapy robot pet would need to recognize true expressed emotional touch in the absence of gestural or emotional direction, contrary to previous classification tasks [4, 43] where participants were asked to *convey* emotion as a proxy to personal experience. Therein lay our first challenge: collecting this kind of truly experienced emotion in lab as training data. We turned to an interesting emotion recall technique. Relived emotional memories can elicit strong biometric responses [32] reminiscent of the original emotional experience. Thus we designed our study around recall of emotionally intense memories and captured this affect-imbued data using our custom touch sensor, supported by commercially available gaze and biometric sensors as well as self-reports.

Abstract

Efficient, unobtrusive machine recognition of human affect will be a key component in interactive systems that must respond to human emotional state - e.g., , robot therapy, assisted-driving systems and emotion-aware game development. Because affective communication occurs through many modalities, algorithmic recognition of affect requires the flexibility to sense and integrate information from multiple sources. Considering the application of therapeutic interaction with a robot pet, we look here at alternative nonverbal modalities known to reflect affect: mainly touch (measuring force magnitude and location), supported by gaze (location) and biometric indicators (skin conductance, blood volume pulse, respiration). We collected a training data series (N=30) from all three modalities and looked for emotion-reflecting features and linkages between them.

For training data that reflected true experienced emotion, we asked participants to relive intense emotional memories while touching a stationary, furry robot, eliciting authentically experienced emotions under controlled lab conditions. Touch data was collected using a sensor embedded under the robot's fur; gaze data was recorded using an eye tracker next to the robot; and biometric data tracked using sensors attached to the participant's body. Each participant interacted with the robot while experiencing two opposing emotions: *Stressed-Relaxed* or *Depressed-Excited*.

Targeting improved emotion classification from integrating touch with gaze and biometrics, we extended past touch classifiers to include features from the frequency domain. We report accuracy results from a random forest classifier built on 100 trees with 20-fold cross validation and leave-one-out using Weka, an open-source machine learning program. Partitioning our dataset by level of system knowledge of participant information [labelled, unlabelled, leave-out] and across time-varying windows [2s, 1s, 0.5s, 0.2s], we found classification accuracy rates improve with increasing system knowledge of participant. Labelled participant data on adjacent windows achieved accuracy rates as high as 92% on touch data alone and unlabelled, untrained participant data as low as 27% (chance 25%).

The wide range in classification accuracy suggests possibilities for *in situ* implementation for a known user base. Adding modalities of gaze and biometric data improve accuracy only when a participant's instances are present in both training

and test sets. To address the computational efficiency required of dynamic and adaptive real-time interactions, we analyze relative subsets of our multimodal feature set, and provide design recommendations for our therapeutic robot.

3.1 Introduction

Social interfaces such as robots, smart cars or game systems must facilitate complex and believable interactions with human users such that the machines appear to respond or adapt to human social cues [34]. Because people prefer to interact with machines as they do with other people [34], it is important for systems to understand human social cues that can carry emotional significance, including nonverbal channels such as facial expression, body pose, social touch, eye focus and vocal prosody. Humans have evolved or learned naturally during early social development to use such cues to distinguish between emotions such as distress and happiness [1]. Machines must be explicitly trained to do this in different social contexts where the affect-expressing channels may vary [1]. Examples of applications with different social contexts are given below:

Entertainment: Emotion manipulation plays an important role in how we experience computer gaming media, but player responses are individual and vary over time. A system able to, for example, detect *excitement* and *startled* could enhance the experience for players who enjoy being challenged or scared, through automatic, personalized changes in difficulty. Technology for detecting emotional state could be embedded in a gaming controller and TV screen.

Assisted driving: In assisted driving it is currently difficult to determine when the system can take over driving safely. Assessing a user's momentary *attention* and *stress* could be a key to this problem. If the car could sense increases in *arousal*, it might control the environment to calm the user down if needed. Sensors could be built into the steering wheel, mirrors, and seat to assess emotional state and focus.

Social robot therapy: Affective therapies for treatment of *anxiety* require systems than can sense human social cues. A number of touchable robots such as the baby harp seal Paro [102] and teddy-bear-like Huggable [95] have been developed for this purpose. Studies with the Haptic Creature [110] specifically investigated human affective touch of and affective display by a zoomorphic robot animal that had

an array of touch sensors embedded in it.

In this paper, we focus on enabling emotion recognition in the last application example – social touch robots for therapy purposes. Earlier work indicates that interaction with such robots can affect human emotional state. [88] showed that motion of the Haptic Creature lowered *anxiety* in users who were stroking it on their laps. However, Sefidgar and MacLean [88] did not set out to investigate automatic human affect recognition with the robot. This could be a key to providing benefits comparable to the physiological benefits demonstrated by animal-assisted therapy [8, 10, 71, 81] – especially valuable where human patients are unable to engage with actual therapy animals.

Using touch as the primary interaction modality leverages the natural inclinations for physical contact to represent emotional closeness while minimizing invasive sensing. Previous work has shown that affect-related information of humananimal robot interaction can be extracted from touch gestures such as stroking and rubbing [4]. However, there is doubt as to whether classifying gestures is helpful in detecting true emotional state – intuitively, knowing whether a gesture was a 'stroke' or a 'rub' may not supply deterministic information about the emotional state of the user while performing that gesture. Furthermore, these studies collected *intent* style data, where the emotions were *expressed to* a sensor, and not *experienced by* a participant. A therapy robot should recognize a user's emotion as it unfolds, thus the model must be built on participants who are truly experiencing the emotions being studied.

Because recognition of human affect from touch data alone is a challenging task, we included two alternative supporting modalities that could potentially improve recognition performance: biometrics and gaze. Our choice of these two modalities over other alternatives such as vocal prosody and facial expressions was motivated through suitability to our robot pet application. Earlier work has shown that touch interaction with a robot pet can decrease heart and respiration rates [88]. This suggests that sensing and utilizing such biometric responses during interaction could make recognition of human affective state more accurate. Additionally, gaze is a good indicator of visual attention determining whether a user is focusing on the robot pet and has been used successfully in predicting affect. Jaques et al [48] has demonstrated that user's gaze focal points on a computer display are related to

feelings of curiosity and boredom.

3.1.1 Approach and Research Questions

We investigate how gaze and biometrics could support recognition of affect through touch interaction with a robot pet. With touch as the central modality, we also chose analysis methods that originate from social touch gesture classification [33, 51] and borrow features calculated from force magnitude and location (pressure-location domain) for finding a baseline for emotion classification in *touch* as well as pressure characteristics in the frequency domain. By adding established signal processing from biometric and gaze classification methods, we hope to reveal how the emotion classification is affected by inclusion of *gaze* and *biometrics*.

Finding an optimal machine learning feature subset becomes combinatorially intractable with increasing modalities and accompanying statistical features¹. However, this is an important endeavour for the eventual end-goal of the creation of a real-time, automatic emotion modelling system. So we must consider data collection window size, sampling density, and feature set to reduce computational load and classification error². Towards this goal, we investigate three avenues: (1) various collection windows and (2) sample adjacency by accuracy, and (3) feature subset performance by selection popularity. Each of these variables are explained in more detail with their respective research questions.

RQ1 – Touch with Multimodal Support: *How accurately can we classify emotional state based on combinations of gaze and physiological data with touch data?*

Multimodal datasets likely provide a more complete picture than touch alone, due to asynchronous activation, or interaction information.

We expect classification accuracy to improve with increased modality

¹Note that our current method of feature creation grows roughly in O(|F||M|), where F is the set of statistical features and M is the set of modalities.

²For a random forest classifier, having extra features that contribute little information can decrease accuracy rates if they are randomly selected. We have also found that certain features oppose each other, i.e. including both in a subset decreases accuracy, when each has high information gain individually.

support. Gaze and biometric data, both known to encode affective content [48, 57], help to round out emotional signals from touch.

RQ2 – System Knowledge of Individual: *How important is system calibration of user to affect classification?*

Past social touch gesture recognition results suggest that individuals have distinctive ways of interacting with touch sensors that make recognizing *identity* surprisingly accurate [19, 33]. This suggests that a system that has learned user behaviour may be better at gesture recognition. Leveraging this result for affect, we perform classification across three different levels of system knowledge of participant (hereby referred to as *participant knowledge*) and discuss results.

Recognition rates is likely to increase alongside greater participant knowledge: Participant-labelled data where instances from the same individual are in both training and test sets will yield the highest classification accuracy; alternatively, lowest predictive accuracy occurs where testing and training are performed on different individuals.

RQ3 – Sample Density: Is classification robust to interruptions in signal or sample size in continuous sampling?

Outside of polling rate, we define *sample density* across two parameters: (1) window size and (2) window proximity. Window size represents a time interval of continuously sampled data; larger windows cover a longer time snapshot and are thus more likely to capture distinguishing emotional characteristics over a behaviour. Window proximity refers to the increased likelihood of neighbouring time series samples sharing more characteristics than distant samples. We examine the influence of window size and window proximity by aggregating data instances in four different window sizes and comparing classification accuracy of the same data set, *with gap* (dropping 2s of data between windows so adjacent windows are not evaluated) and *without gap* data (adjacent windows are included in the training and test sets).

We posit that across both parameters, reducing sample density reduces

classification accuracy where the worst performance occurs with small windows with gapped data.

RQ4 – Feature Analysis: Which features or sets of features provide the best chance of high accuracy rates?

Increased computation load and resultant latency are the cost of multimodallyenabled triangulation on parameter truth, and potentially undermine realtime feasibility of such a technique. In order to optimize these tradeoffs, we analyse each of our features in terms of repeated occurrence in automaticallyselected best feature subset. Finally, we assess which features, from a superset containing both *pressure-location domain* and *frequency domain*, are to be included in a strong feature set.

Traditional touch and gaze features span the spatial and temporal domain, we add spectral distribution statistics to incorporate artifacts in interaction frequency and posit that **classification accuracy will be improved with the addition of** *frequency domain* **features**.

3.1.2 Contributions

Through answering our research questions, we contribute the following:

- Affect classification performance from combinations of touch, gaze, physiology data in *experienced-emotion* interactions;
- *Feature analysis* distinguishing relative recognition contributions from feature subsets, to maximize multimodal benefits tradeoff with computational load as well as a recommendation of data features from the frequency domain and traditional pressure-location domain in an emotion classification context;
- Practical recommendations on the use of affect classification in three example scenarios with respect to system knowledge of the user.

The remainder of this paper is structured to include a survey of previous work, motivating our choice of emotion elicitation method and including a history of affect classification from each of touch, gaze, and biometrics. We describe our experiment in detail, describing our data post-processing procedures. Results are presented from all data partitions that target the influence of multimodal data vs touch alone, participant knowledge, sample density, feature set, as well as a review of the emotional experience based on participant reports. Finally, we discuss our findings and ground this work in relevant application implications.

3.2 Related Work

Here we review past methodologies and motivate the choice of relived experience to target our emotion set. We also provide a history of affect classification by modality and note cautions.

3.2.1 Emotion Set

We take Russell's circumplex model of affect to be our starting point, where arousal (activation) and valence (pleasantness) are orthogonal axes [83]. Although Russell's model is widely used by emotion researchers, there are some inherent limitations with discretizing and labeling the two-dimensional space as we must assume that: (1) emotion labels will be interpreted consistently by every participant at any time; and (2) the axes are truly orthogonal.

Consider the emotional context of approaching the axes or origin when working with such a model:

- it may not be sensible to reach (0,0), presumably a state of full neutrality. Similarly,
- it may be absurd to talk about independent movement, i.e., , directly along axes: can one really have an increase in arousal without any change in valence?

As such, many emotion researchers [27, 42, 105] opt to discretize the 2D space into a grid and rotate it by 45, such that experimental materials and tasks are aligned with the diagonal axes, namely (high arousal, high valence) \leftrightarrow (low arousal, low valence) and (high arousal, low valence) \leftrightarrow (low arousal, high valence). The literature provides little consensus for which emotion labels are to be employed, making comparison between studies of even common modalities problematic. Understandably, papers utilizing information of gaze use attention-related emotion sets – e.g., , *Anxiety, Boredom, Confusion, Curiosity, Excitement, Focus, Frustration* [84]. Human recognition of human affect based on touch tries to span the human experience, namely *Anger, Fear, Happiness, Sadness, Disgust, Surprise, Embarrassment, Envy, Pride* [43]. Yet another method is to partition Russell's affect grid as discrete labels: touch uses nine³, of which biometric data uses a subset⁴. We chose four emotions labels that minimally cover the Russell's emotion model to avoid overlap in label interpretation: *Stressed, Relaxed, Exited,* and *Depressed*.

3.2.2 Emotion Elicitation

A fundamental problem with eliciting emotions from research participants [22] is consistently producing *valid* emotions in a lab setting. Our motivating application involves a situated social robot that must react to authentic human emotions as they occur, which poses a challenge in contrived laboratory settings (*"Feel angry as you interact with our robot"*). To circumvent these artificial emotion barriers, past work has typically asked of participants the easier task of simulating intended emotions (*"Imagine feeling anger, then express it to our robot"*). For example, to collect the data used in [110] and [4], participants were presented with a list of emotions that they had to express by touching a robot. Unfortunately, this does not directly equate to *experiencing* an emotion, and may not accurately represent the intensity of an experienced emotion. Consider the difference between a smile and genuinely feeling happiness: the former lives in a public space, where the one smiling can convey an *emotional intent* to others. The latter is private, the *experienced emotion* available only to the individual.

Experienced emotion studies are difficult to conduct and are more rare. Playing entertainment media, specifically emotionally evocative music and/or video, has been employed as an emotion elicitation method [57]; however, selecting the

³Emotions for classification by touch: *Distressed, Aroused, Excited, Miserable, Neutral, Pleased, Depressed, Sleepy, Relaxed* [4].

⁴Emotions for classification by biometrics: *Stressed, Excited, Depressed, Relaxed* [57].

media to be applied as an emotion treatment introduces difficulties. Although validated sets of emotional media exist, they either rely heavily on getting the mood 'just right' (music), or demand a high level of attention and divert gaze (video). Further, cultural and individual differences have great and unpredictable influence over emotional reactions to either medium.

To channel our scenario of interacting with a social touch robot, we asked participants to recall and/or retell an emotionally intense memory (as supported by [32, 62]) while interacting with a furry, touch-friendly robot. Specifically, participants were asked to recollect a memory where they had strongly experienced a specific emotion, and recall or tell the story of that experience to the robot while touching it. To verify that participants felt the intended emotional state, we used a self-report scale where they rated their current feeling before and after a task.

3.2.3 Modalities

Affect classification has been explored in each of the modalities explored here, all with distinct emotion elicitation procedures and labels.

Touch

Touch data can be quickly dissected into force magnitude (or pressure) and location – dimensions which are used for gesture recognition as well as for control directives (e.g., , using trackpads and touch screens). Social touch gestures have been studied and prediction accuracy has ranged from 59% [51] to 86% [33] depending on collection and classification methods, which, like affect, have no consistent standard. Still, the high prediction accuracy rates achieved on defined gesture sets suggest that these touches can be used as directives in systems with embedded recognition systems.

Accurate recognition for true emotional data analysis seems to be more difficult: even human recognition of human emotion in touch does not achieve gesture recognition rates with highs of 59% (chance 8%) [43]. Machine classification has demonstrated $36{\sim}48\%$ accuracy (chance 11%) [4] depending on inclusion of participant knowledge. Both of these studies collected *intent* data, and not *experienced* relived emotion.

Gaze

An interactant's eyes give affect cues which are discernible with eye tracking technology. Partala and Surakka [74] studied the effect of emotional auditory stimulation on pupil size variations; they found that negative and positive stimulation resulted in significantly larger pupil dilation than neutral stimulation but could not differentiate stimulus valence. Other factors, such as changes in luminance, can also affect pupil dilation.

An alternative is to analyze where a person is looking. Jaques et al [48] tracked students' gaze when they interacted with a graphical intelligent tutoring system; gaze features such as fixations and saccades revealed that curious and bored students looked at different interface areas – for example, engaged students looked more at the table of contents. Overall, boredom and curiosity could be predicted with 69% and 73% accuracy, respectively.

To our knowledge, no studies have investigated whether gaze *point* is useful in classifying emotions using the two-dimensional model of valence and arousal. Compared to pupil size variation measurements, gaze point can be measured in a less controlled environment (lighting and luminance changes impact data quality less) with relatively inexpensive tracking technology. Thus, we utilize the Cartesian coordinates of user gaze point in our own classification analyses.

Biometrics

Biometric signals such as blood volume pulse (BVP), skin conductivity (SC) and respiratory rate (RR) have been widely used for multimodal emotion recognition in a variety of contexts. Examples include facial expressions [60], affective audio [57, 69], gaze behaviours [45], and touch behaviours [88]. In addition, heart rate variability has been utilized in emotion classification [5, 6, 49, 88] and as a biophysical indicator of cardiological health [76].

Our work follows in a tradition of earlier explorations into multimodal emotion recognition. The biometric signals that we chose were blood volume pulse (BVP), skin conductivity (SC), and respiratory rate (RR). Using these three basic signals, we calculated a set of derived signals that consisted of heart rate variability (HRV) features, breathing rate variability (BRV) features, and cross-signal features such as heart beats per breath. Studies where emotion elicitation is based in true experience and uses the same emotion sets (as in [57]) are most appropriate for comparison. [57] uses validated music excerpts to generate authentic responses crossing four musical emotions (positive/high arousal, negative/high arousal, negative/low arousal, positive/low arousal), reporting affect recognition rates between 70% and 95% (chance 25%) [57], with higher rates where participant knowledge is included.

3.3 Methods

We asked participants to recall emotionally intense experiences while interacting with our stationary robot pet. We sampled touch, gaze and biometrics sensors, with self-reports of emotion collected before and after each emotion. Of 30 participants recruited from across campus, 14 identified as female, 18 had corrected vision, and of mean age 25.4 years (σ =5.4 years). Participants were compensated \$20 for their time.

In the following subsections we give details of the experimental setup, introduce the procedure, and describe our data post-processing techniques.

3.3.1 Experimental Setup

Configuration and Room

Participants were positioned in a half-prone position on a couch to reduce largescale movements while ensuring comfort (Figure 3.1). For valid data collection, our gaze and touch tracking systems needed to be within a close range; for valid emotion elicitation, participants had to be comfortable enough during memory recall to express the focal emotion. The experiment was conducted in a sparsely furnished medium-sized office with a window, with the participant's back to the door. The experimenter was present except for emotionally intense parts of the session, as described below.



Figure 3.1: The experiment setup: participant sits comfortably supported by pillows, facing the gaze tracker with her hand on the touch-sensitive surface of a stationary robot. Biometric sensors are worn around the waist (respiratory rate), thumb (blood-volume pulse), and index and ring fingers (skin conductance) of the resting hand. One camera captures eye movements and another is raised on a tripod behind the participant to capture hand motions over the robot. Both cameras have audio disabled for privacy. When the participant pulls the rope, a ball leading outside the room indicates to experimenters that the emotion task is complete.

Touch Sensor on a Passive Robot

We used a custom flexible touch sensing apparatus previously described in [19] that has been validated to detect $5g\sim1kg$ of weight with resolution of 10×10 inches at one taxel per square inch⁵. We chose fingerpad-size taxel resolution similar to that of earlier work [4, 33] since emotion tasks in touch generally incite broader movements [43]. Higher resolution sensors have had great success in high precision tasks such as those seen in touch screens, trackpads, or teleoperative mimicry and have useful applications in robotic arms [93]; however, they are massively overqualified in terms of computation load and resolution for our purposes, where low cost and high sensor malleability are crucial.

Forming a 10-by-10 grid, this device can sense multiple simultaneous touches (so-called multitouch), registering varying pressures on each taxel scaled to 1024 levels and polling at 54Hz. This resulted in 54 frames of 100 cells per second, each reading a touch pressure value between 0 and 1023.

The touch sensor was installed on a stationary and unresponsive furry object, roughly the size and weight of a football. The sensor was affixed by velcro to a substrate made of soft-shelled binder plastic⁶, then covered with a uniformly-textured short, soft minky fabric⁷ described as "*pleasant to touch...[and] reminded me of my chocolate lab's head*" – P4. Participants were instructed to touch the top surface of the furry robot (Figure 3.1). All sensors were wired through the robot platform to minimize visual clutter and connected to a single laptop. Figure 3.2 demonstrates robot construction. While the robot is capable of motion, we disabled actuation for this study to reduce confounds from novelty effects, sounds, or expectation of a reactive robot.

Gaze and Biometric Sensors

We sampled gaze behaviour via Tobii EyeX gaze tracker⁸, at a rate of 60Hz – similar in rate to our touch collection. The Tobii EyeX tracker was chosen because it has been used successfully in a number of studies to track users gaze location when looking a computer display or tablet [55, 61, 70]. It is also small in size so we could place it below the robot at an angle where it could see participant's eyes while touching the robot (Figure 3.1). Although no specific instructions were given regarding a required gaze direction, participants were informed that gaze data collection was best when they were facing forward and did not make large body movements.

We collected three biometric signals using the Bio-Graph Infiniti Physiology Suite⁹, namely Blood Volume Pulse (BVP), Skin Conductance (SC), and Respira-

⁵Built from commercially available piezoresistive and conductive fabric. Fabric is commercially available at www.eeonyx.com.

⁶Specifically, Wilson-Jones Accohide flexible cover binder http://amzn.to/29AiYbx.

⁷Minky is a soft, short-pile fabric, often used to make baby blankets. Examples can be found online including http://amzn.to/29L2DBv.

⁸Tobii EyeX gaze trackers are available at http://www.tobii.com/xperience/products/.



Figure 3.2: Our robot was constructed from pliant plastic sheets actuated by a pulley, covered with a custom-built sensor, and finally wrapped in a furry fabric to invite touch. It remained stationary during the course of this study to eliminate any effect of participant reaction to robot motion.

tory Rate (RR), all collected at 2048Hz. Following established procedures for derived signals, these were expanded to include a set of Heart Rate Variability (HRV) features, Breathing Rate Variability (BRV), and cross-signal features such as heart beats per breath. The cross-signal features were automatically calculated as part of the Thought Technology Physiology Suite.

Participants were first outfitted with a respiration band worn around the chest

 $^{^9}$ System manufactured by Thought Technology Ltd. FlexComp ∞ SA7550 Hardware Manual can be found through manufacturer website at http://bit.ly/29A5NIC.

- a close fit that did not impede breathing. Once they were comfortably seated, we positioned the BVP sensor at the thumbpad. Finally, the SC sensors were positioned on the index and ring finger pads. Both the BVP and SC sensors were held in place by a small velcro band on the right hand which was not used for touching the robot.

Video Data Collection

We took video recordings of the participant's hands and face to help in data analysis in case of missing gaze or touch data. No sound was recorded out of respect for privacy. Camera for recording the hands was placed behind the participant and another camera was placed on the right side of the participant for recording the face (see Figure 3.1 for placement of the gaze camera).

3.3.2 Procedure

Table 3.1 summarizes the study procedure. The main part of the experiment was run in alternating stages: (1) a neutral task, (2) the first emotional task, (3) a neutral task, (4) the second emotional task. Emotion tasks were counterbalanced across participants.

Introduction and Calibration

The participant was welcomed, consent process administered, and sensing equipment set up. To reduce novelty effects, we introduced the robot, invited touch exploration, and described the construction of the robot including sensing abilities. We explicitly indicated that the robot would not move throughout the study. Data reading range was then checked for each sensing modality in a calibration phase.

Neutralization and Self-report

For each stage, we first presented an emotionally neutralizing reading task where the participant read a short technical report from a technology magazine. The participant then reported their current emotional state by marking a sheet of paper displaying Russell's [83] 2-dimensional scale of arousal and valence [22]. During this period, the participant was instructed to keep their hands still and not touch

Procedure	Description (imposed duration)	Data Recorded		
Introduction	Describe experiment procedure	none		
	Apply sensors	none		
	Verified stable Biometric readings	none		
Neutralization 1	Read neutral text (5 min)	Biometrics		
	Self-report	Emotional State		
	Calibrate gaze tracker and touch sensor	Calibration Logs		
Emotion 1	Recall Memory	Biometrics, Gaze, Touch		
	Self-report	Emotional State		
Neutralization 2	Read neutral text (5 min)	Biometrics		
	Self-report	Emotional State		
	Calibrate gaze tracker and touch sensor	Calibration Logs		
Emotion 2	Recall Memory	Biometrics, Gaze, Touch		
	Self-report	Emotional State		
Gesture	Calibrate gaze tracker and touch sensor	Calibration Logs		
	8 randomized gesture tasks (10s each)	Gaze, Touch		
Debrief and Interview	Interview	Qualitative data		
	Self-report	Emotional State		

Table 3.1: Experimental Procedure Summary

the robot. A definition of arousal and valence were provided, and an experimenter answered any and all participant questions about the nature of the scale. This self-report was repeated before and after each neutralizing and emotion task. For emotion tasks, a genuineness rating of their experienced emotion was included.

Reliving Emotion Task

The participant was next instructed to recall an emotionally intense memory pertaining to a given emotion word {*Stressed*, *Excited*, *Relaxed*, or *Depressed*} and, while comfortably seated, interact with the furry sensing robot. They were instructed to relive the emotion as intensely as possible and describe the memory and associated feelings to the robot in any language they were most comfortable in at a volume of their choosing. We reminded them that all audio recording in video cameras was disabled and that we could not hear them speak from outside the experiment room. Task completion was indicated by pulling a rope that led to where experimenters were waiting. Data was collected over the course of a single recalled memory (duration μ =4.23 min, σ =3.09 min). Once the participant pulled on a rope to indicate that he or she had completed the emotion task, the experimenter returned and administered the self-report grid.

The participant then proceeded to the second set of neutralization and reliving emotion tasks. The emotion of the second emotion task was determined by the first; participants experienced, in counterbalanced order, either *Stressed - Relaxed* OR *Depressed - Excited*. Due to the taxing nature of the emotion task, we chose to use only two emotions per participant, determined through discussions with field experts, piloting, and literature review. Since feeling genuine emotion can take effort, there was a significant concern about fatigue effects. The condition sets for the neutralization and emotion tasks are fully summarized in Table 3.2.

Gesture Task

To ensure touch data quality, participants then performed a series of nine touch gestures for 10s per gesture on the touch sensor. This gesture set was chosen from Yohanan et al.'s Social Touch Dictionary [110] and is consistent with the tasks as published in Cang et al. [19] using definitions as described in Table 3.3.

Task	Description	Order	N=30
Neutral	Two neutral texts SP1 P2	(R1, R2)	16
	Two neutral texts $\{\mathbf{R}^{T}, \mathbf{R}^{T}\}$	(R2, R1)	14
Emotion	- 4 Emotion labels [83]	Stressed -	6
	- Each participant recalled	Relaxed	
	"opposing" emotions	Relaxed -	8
	- Counter-balanced by order	Stressed	
	-	Depressed	8
		- Excited	
		Excited -	8
		De-	
		pressed	

Table 3.2: Summary of the condition sets for Neutralization and Emotiontasks. For example, one participant session consisted of R1 neutral task,*Stressed* emotion task, R2 neutral task, and *Relaxed* emotion task.

Table 3.3: Touch gesture instructions as provided to participants [19].

Gesture	Definition Provided
no touch	no contact with the sensor (control)
constant	touch contact without movement
pat	quick & gentle touches with the flat of the hand
rub	moving the hand to and fro with firm pressure
scratch	rubbing with the fingertips
stroke	moving hand repeatedly
tickle	touching with light finger movements

Debrief and Interview

Finally, we conducted a short interview and debrief, including a final self-report to ensure that the participant felt comfortable and emotionally stable. The latter was included after we found in piloting that participants could become very distraught during these sessions. The entire experiment took approximately 60 minutes.

Data Processing and Feature Extraction

To check against sensor and/or data degradation, touch gesture data collected during the Gesture Task as described in Table 2.1 was processed identically as that in [19] where the same touch sensor was attached to a slightly larger robot¹⁰.

The results of earlier work were consistent with the current gesture data results showing that gestures such as stroke and rub could be differentiated also with the touch sensor mounted on a smaller robot. Thus, the touch sensor data quality should be sufficient for classification of affect-related touch data in the current study.

The recorded touch, gaze and biometric data was used to calculate features that could be used in affect classification. We included conventional touch features [4, 19, 33]: min, max, mean, median, variance, total variance, area under the curve for both pressure and touch centroid (x,y). Then we extend these features to gaze focal location (x,y) and biometric channels of blood volume pulse (for heart rate), skin conductance, and respiration rate. As there may be more information than magnitude and direction in touch and gaze behaviour [4], we also extracted frequency-domain features, specifically fundamental frequency, amplitude, and peak count for touch: frame-level pressure, frame-level centroid (x,y), and the window's nine most visited cells as traced by the centroid; and for gaze: point of focus on surface $(x,y)^{11}$.

We down-sampled biometric data from the original 2048Hz to match touch and gaze sampling rates at 54Hz–60Hz. All features were then calculated over different window sizes: 2s, 1s, 0.5s, 0.2s, both with and without gaps (See Figure 3.3).

Pressure-location Domain Emotional touch feature extraction reprises known analysis procedures for social touch gesture recognition constructing three pressure parameters: {framesum (sum of all taxels in each frame), row-centroid and column-centroid (weighted measurement of centre of mass)} [19, 33]. These parameters are split into windows, where a 2s window contains 2s of data, or 106 frames (54Hz

¹⁰For our purposes, the larger robot was not feasible, since it had poor sensor coverage on key parts of its body, and was not able to be easily fixed in relation to the eye tracker.

¹¹Biometric channels also included automatically calculated higher-order frequency features that came pre-packed with the Thought Technology physiology suite, so further frequency calculations on biometric data was not reasonable.



Figure 3.3: The data recorded during a single emotion task is referred to as a *capture*. To determine the effect of varying window size on accuracy rates, we tested at 2s, 1s, 0.5s, and 0.2s windows. Assuming that there may be similarities learned when trained and tested on directly adjacent windows, we also compared accuracy rates of data with and without imposed gaps to remove adjacent window instances.

 \times 2s – Figure 3.3). Seven statistical features (min, max, mean, median, variance, total variance, area under the curve) were calculated for each parameter, for a total of 21 touch features.

Gaze used a total of 34 features: mean, min, max, median, variance of each of focal coordinate pair (X-, Y-location), saccade length, velocity, fixation duration as well as window summary features of sample count, sample count on robot, samples off robot, sample ratio, within range rate, saccade count, saccade rate, fixation count, fixation-saccade ratio. We used Salvucci's I-VT algorithm [87] to differentiate between fixations and saccades. Gaze samples with point-to-point velocities lower than 30 degrees/second were classified as fixations. Samples with velocities
of 30 degrees/second or higher were classified as saccades.

Biometric features included mean, median, and variance across all channels provided from the Thought Technology physiology suite, including both base signals (specifically BVP, SC, RR), and higher order channels dependent on the original signals (e.g., , HR, HRV, IBI, etc.), for a total of 228 features across 76 channels.

Across all three modalities of touch, gaze, and biometrics, our maximal pressurelocation feature set contained a total of 283 statistical features.

Frequency Domain As biometric features extend far beyond these simple statistics, we elected to compare frequency-domain features for touch alone and touch and gaze interaction.

Inspired by [4]'s use of *max amplitude* and *most activated cell* features in the frequency domain, we extended the set. From the original touch dataset, we separated each collection instance to various windows and from each window we performed a Fast Fourier Transform (FFT) of the frame-level pressure and the centroid coordinates (x,y). Tracing the centroid, we found each frame-centroid cell as well as its eight surrounding cells and the FFT for each of these nine cells. From these combined 12 parameters, we found six window statistics: the spectral peak count, fundamental frequency (Hz), max amplitude, mean amplitude, variance and total variance of amplitude, resulting in a total of 72 touch features in the frequency-domain.

We extended the same six window stats to gaze data and on the 2D focal location, calculating 12 gaze features.

Across touch and gaze, our maximal frequency domain feature set thus contained 84 statistical features. When we include the 21 touch pressure-location set to assess all touch features, this number increases to 105 statistical features.

Classification Consistent with earlier work in touch classification [4, 19, 33], we use Weka, an open-source machine learning application [41], to run 20-fold cross-validation (CV) using a random forest classifier of 100 trees to report results in terms of classification accuracy on both the pressure-location and frequency domain features. Whether for 20-fold CV or Leave-One-Out, classification accuracy is defined as ratio of correctly classified instances over all instances. We use random forest as it has shown to be effective in touch classification tasks in past

Table 3.4: A motivating overview of analysis techniques.

Description	The data is randomly partitioned into 20 equal sized blocks (folds). One fold is held as the test set, and the other 19 folds are used as the training set. 20 tests are run, one for each fold, and results are averaged.
Implication	Since the data is randomly partitioned, all participants are likely to be in both the test and training sets.
Question	How can we expect a system with prior knowledge of participants to perform?
	(b) Leave-one out

(a) 20-fold cross-validation

Description	One full set of samples are left out of the training set and kept as the training set. Since each participant did two emotions, we left two participants out for each LOO test. Results are generated per left out pair.
Implication	Since the data is systematically partitioned, two participants will be completely left out of the training set for each run.
Question	How can we expect a system with no prior knowledge of participants to perform?

studies [4, 19, 33, 51].

A summary of analysis techniques is reviewed in Table 3.4.

Varying window sizes informs real-time classification in case we choose to limit how much data to collect before a prediction task. Windows for capturing gesture data have been recommended at 2s [33]; however, human hand and finger expression is much quicker at roughly 100-200ms [72, 92].

Therefore, we divided our data into increasing window sizes of [0.2s, 0.5s, 1s, 2s], to provide us with insight into how data length influences classification performance. Larger window sizes reduces the interrupts required and thus overall computation time. We also assess the influence of removing adjacent windows by adding 2s gaps between classification instances and comparing performance when

all windows are included.

In decreasing knowledge levels, we *included* participant labels, *excluded* participant labels, and finally used a version of Leave-One-Out (LOO) where wherein we trained a classification model on a dataset with test participants removed. In other words, if we are classifying Participant 3's data instances, Participant 3 is absent in the training set and only appears in the test set. To ensure that we covered all four emotions in each run, test sets were comprised of two complementary participants. Since each participant performed two emotions, we had two mutually exclusive groups of emotions. Combining one from each group gives us 16×14 combinations or 224 unique test and training sets for each modality set and window size. We aggregated for modality combination and window size and report classification accuracy.

The emotion classification tasks varying participant knowledge can be thought to represent real-world applications:

20-fold CV was done *with* and *without participant labels*. Training the model *with participant labels* simulates a system that knows whose emotions it is attempting to classify, i.e. has some *a priori* knowledge of the user.

20-fold CV *without participant labels* can be read as a simulation of a classification task where the interactive system's emotion model has been trained on all possible users before attempting classification, though not given explicit indication of the current user. Imagine a system that lives in a limited private domain, such as family or classroom context, where all users have gone through a calibration period. The calibration data would be the training set for the model.

In contrast, the *Leave-one-out analysis* can be read as a simulation of a classification task where the interactive system's emotion model cannot be trained on all possible users. This could work for a system that lives in a public domain, such as in a museum or institutional context. The training set for the model would necessarily not include all possible users, since they would not be known ahead of time.

We also ran 20-fold CV classifying on participant labels to determine not only how well these feature sets can determine *what interaction* was performed, but also *who* performed it.
 Table 3.5: A motivating overview of analysis conditions.

(a) Window size

Description	Data was all sampled to 54Hz. Window size is the length of time over which a feature is calculated. E.g. a two-second window has 108 samples.
Implication	With a static sample speed, shorter windows simulate a system with a faster update cycle. There will be less information per window, but a faster system response.
Question	How can we expect accuracy rates to change with different sample sizes?
	(b) Gap/without gap

Description	With no gap, all windows are calculated contiguously, i.e, every window is directly adjacent to the one previous. With gap, after every window is calculated, two seconds of data is thrown out.
Implication	Previous work has shown that a touch gesture takes a little under a second to make, so a two second gap increases the likelihood that each window captures different gestures.
Question	How does a system trained on more homo/heterogenous data instances perform?
	(-) D $(-)$

(c) Participant labels in/out

Description	With participant in, participant labels were one of the features the system could randomly select.
Implication	With participant in, the system can tell whose emotions it is attempting to predict. With participant out, the system still has knowledge of the participant's behaviour, but cannot tell from whom the behaviour came.
Question	If the system is calibrated on a set of individuals, does it need <i>a priori</i> identification of an individual to predict emotions well?

3.4 Results

This section describes our results from running classification using the traditional pressure-location touch features on a multimodal data set as well as our addition of Frequency Domain features on touch and gaze data. For both feature sets, we found similar patterns, where:

- Increasing number of modalities improved accuracy rates,
- Increasing window size had little effect on accuracy rates,
- Decreasing amount of participant information in the training set **worsened** accuracy rates,
- Leave-one-out analysis performed at or near chance, and
- Participant classification performed comparably to emotion classification.

In each case, we return to our research questions and describe the feature subset run on emotion and participant and the nuances from the result set. We also address the performance of feature set analysis, reporting best performers for emotion classification. Finally, we detail participant's self-report of their personal experiences during the emotion expression tasks.

3.4.1 Pressure-location Feature Set: Emotion Classification

Pressure-location touch features have performed well in earlier touch gesture recognition tasks [19, 33]. To compare the relative classification accuracy with frequency features, we report results from random forest classifiers using 20-fold cross-validation (CV) on varying degrees of system knowledge by four differing window sizes. We aggregate across combinations of modalities: touch-only, touch/gaze, touch/biometrics, and touch/gaze/biometrics, reporting on the relative impact of each modality set. Biometric classification accuracies are included both for reference to accuracies found in past work [57], and as a check against the performance of touch and gaze features.

Table 3.6 is separated into two sets of classification tasks where (1) instances were adjacent therefore likely to be similar and (2) instances were separated by 2s



Figure 3.4: Classification performance by modality set, level of system knowledge (where *labels in* indicates with knowledge), window size, and with/without gaps. The **Emotion** row displays results from classifying emotion, with and without **Participant** labels. The **Participant** row displays results from classifying participant, with and without **Emotion** labels. For both cases, increasing the number of modalities improves accuracy rates, where the inclusion of biometrics provides the strongest classification rates. However, in the most rigorous test of emotion classification, Leave-one-out (LOO), all modalities perform at or near chance for all window sizes. Regardless of classification task, window size has a small or positive effect on accuracy rates, except for case of 2s-windows-with-gap. This is likely due to a high decimation of data, i.e. 2s-windows-with-gap.

Table 3.6: Results from Classifying Emotion using 20 fold CV on pressurelocation features for Touch (T), Touch + Gaze (T+G), and Touch + Gaze + Biometrics (T+G+B)

Window	Т	T+G	T+B	T+G+B	Count	
0.2s	92.29	94.41	100	100	73835	
0.5s	92.14	93.94	100	100	29950	
1s	92.13	93.71	100	100	14995	
2s	91.84	93.05	99.96	99.97	7435	
(h) With 2s Gan between instances participant labels in						

(a) No Gap between instances, participant labels in

(0)	With 25	Gap	JULWU	JII IIIS	stanc	cs, pa	intici	pan	ii iai	JCIS III	
	I		1 _	~		_ 1		~	_	~	

Window	Т	T+G	T+B	T+G+B	Count
0.2s	87.13	88.41	99.99	99.97	6713
0.5s	88.15	89.68	99.93	99.98	5990
1s	88.56	90.96	99.9	99.92	4999
2s	76.33	76.48	93.79	93.05	676
$(a) \mathbf{N} \mathbf{a} \mathbf{C}$. :			, 1

(c) No Gap between instances, participant labels excluded

Window	T	T+G	T+B	T+G+B	Count
0.2s	75.42	82.2	99.99	100	73835
0.5s	76.11	82.35	99.99	99.99	29950
1s	76.77	82.15	99.98	99.97	14995
2s	76.15	81.45	99.93	99.88	7435

(d) With 2s Gap between instances, participant labels excluded

Window	T	T+G	T+B	T+G+B	Count
0.2s	66.75	70.36	99.87	99.82	6713
0.5s	69.4	73.62	99.77	99.73	5990
1s	71.93	75.88	99.8	99.76	4999
2s	56.66	60.06	89.79	90.09	676
(e) Leave-participant-out Pairs					

Window	Т	T+G	T+B	T+G+B
0.2s	30.62	30.04	30.14	29.78
0.5s	31.95	31.65	27.24	27.86
1s	32.19	33.08	27.83	27.75
2s	33.05	34.47	29.14	28.22

to reduce similarity; results show better performance where instances were adjacent. Gaze and biometric support greatly improves the accuracy rate, in line with previous work showing high classification performance on physiological data [57]. Generally, increasing window sizes provides more information. However, we see a drop-off at 2s. We believe this is due to a reduction in training instances, for example, in Tables 3.6b and 3.6d, the instance count of 2s windows with gap has only ~13% of the number of instances as the next window.

As can be seen in Tables 3.6a–3.6b, we report accuracy rates between 65 and 100%, depending on window size and modality combination. Results are presented in order of decreasing participant information inclusion in the training set i.e., participant labelled, participants unlabelled, leave-participants-out (LOO).

Tables 3.6c–3.6d shows a similar pattern as with participants labelled (above). Note that with lesser participant information, Touch and Touch+Gaze performance degrades; however, the inclusion of biometric data maintains stable accuracy rates.

In our most rigorous test, **LOO**, high classification rates here would indicate generalizable emotional expression of the population. However, this is not evident from our feature set under these experimental conditions as, regardless of modality, results from Table 3.6e approach chance (25%).

The bargraph of accuracy rates from all Leave-Participants-Out test set is shown in Figure 3.5. Mean rates approach chance (25%), suggesting that the average individual's behaviour during emotional tasks is not generalizable. Interestingly, there are many individuals who may conform to the group or directly oppose the group as can be seen in the extrema of the distribution. For example, we note that P22 is exceptional in that when P22 is included in the test set with any other participant, emotion classification was consistently well-above chance.

3.4.2 Frequency Domain Feature Set: Emotion Classification

An earlier study by Altun and MacLean [4] reported promising results using two frequency domain features (peak frequency and corresponding cell index) over the classic pressure-location touch features. Here we replicate and extend their feature set over both touch and gaze modalities. Biometric data was not included here since physiology feature calculations have far more extensive feature sets beyond



Figure 3.5: This is a bargraph of accuracy rates for every LOO test by participant. Most of the accuracy rates are roughly chance=25%, and low SD suggests the variance of this classification was low across window sizes. Notice a few interesting outliers: test sets including P22 and P05. P22 has much higher classification rates, suggesting that P22 may be similar enough to the group to have their emotion behaviours consistently identified. Conversely, P05's particularly low rate of classification suggests that P05 expresses emotions contrary to the group.

these frequency domain calculations [6, 57].

We again calculate features across the same four window sizes and compare the influence of removing adjacent windows. Results are presented based on data from Touch Frequency (TF), Touch Frequency/Touch pressure-location (TF+T), and Touch Frequency/Touch pressure-location/Gaze Frequency (TF+T+GF).

Repeating the same classification tasks from the pressure-location feature set, we compare how frequency features impact recognition of emotion.

With participant labels, we see the highest performance in cases with high similarity and high information: adjacent (without gap) instances of multi-modal, large windows. Without participant labels, the loss of adjacent windows (with gap) has a large negative impact on recognition. Table 3.7b shows that even in the best case (multimodal, largest window), the added gapping results in large accuracy loss. Without participant labels nor test participant data (LOO) results again approach chance (25%), similar to that of the pressure-location feature set.

Table 3.7: Results from Classifying Emotion using 20 fold CV on Touch Frequency (TF); Touch Frequency+Touch pressure-location (TF+T); Touch Frequency+Touch pressure-location+Gaze Frequency (TF+T+GF) feature sets.

TF	TF+T	TF+T+GF
78.04	91.83	91.88
86.48	93.57	93.48
81.67	90.58	90.88
92.98	95.79	96.15
	TF 78.04 86.48 81.67 92.98	TFTF+T78.0491.8386.4893.5781.6790.5892.9895.79

(a) With no gap between instances, participants in

(b) With 2s Gap between instances, participants in

_

Window Size	TF	TF+T	TF+T+GF
0.2s	70.69	83.7	83.9
0.5s	83.1	89.64	90.25
1.0s	76.91	85.74	85.64
2.0s	82.52	89.07	89.49
() W'' (1)		· .	',•• , ,

(c) With no gap between	instances,	participants	out
-------------------------	------------	--------------	-----

Window Size	TF	TF+T	TF+T+GF
0.2s	78.04	76.68	76.96
0.5s	86.48	81.9	81.93
1.0s	81.67	75.35	76.38
2.0s	92.99	90.32	90.38
	'		'

(d) With 2s Gap between instances, participants out

Window Size	TF	TF+T	TF+T+GF
0.2s	42.71	62.36	62.76
0.5s	67.67	76.07	76.86
1.0s	51.08	67.32	68.16
2.0s	58.54	74.17	75.5



Figure 3.6: Feature selection popularity of pressure-location features of Touch and Frequency features of Touch and Gaze.

3.4.3 Participant Classification

The sharp decline in classification accuracy associated with eliminating participant information from training sets in Leave-Out needs further examination. To this end, we performed 20-fold CV predicting on participant for both feature sets (see Table 3.8 for pressure-location features and 3.9 for frequency features).

High accuracy rates on participant prediction for pressure-location features suggests that individual differences are highly expressed in behavioural data examined at this level. We see a greater negative impact on accuracy from removing emotion labels Table 3.8a to 3.8c than removing adjacent instances Table 3.8a to 3.8b. However, compared to Section 3.4.3, frequency features perform weakly for predicting participant.

3.4.4 Feature Set Analysis

Participant information makes a notable difference in classification accuracy of emotion across all window sizes and adjacency. In order to examine how classic touch features of pressure-location compares to an extended Frequency features on Table 3.8: Results from Classifying Participant using 20 fold CV on pressurelocation features for Touch (T), Touch+Gaze (T+G), Touch+Bio (T+B), and Touch+Gaze+Biometrics (T+G+B)

Window Size	Т	T+G	T+B	T+G+B	Count
0.2s	84.66	90.48	100	100	73835
0.5s	85.63	90.94	100	99.99	29950
1s	86.18	91.24	100	100	14995
2s	86.24	91.61	100	99.99	7435

(a) No Gap between instances, emotion labels in

Window Size	Т	T+G	T+B	T+G+B	Count
0.2s	75.57	81.68	99.93	99.96	6713
0.5s	78.88	84.36	99.97	99.95	5990
1s	81.5	87.04	99.94	99.98	4999
2s	67.31	71.3	98.52	98.22	676

(b) With 2s Gap between instances, emotion labels in

Window Size	T	T+G	T+B	T+G+B	Count
0.2s	71.96	81.98	100	100	73835
0.5s	73.64	83.28	99.99	99.99	29950
1s	75.43	84.61	100	100	14995
2s	75.74	84.87	99.96	100	7435
(d) With 2s Gap between instances, emotion labels excluded					

(d)) With $2s$	s Gap	between	instances,	emotion	labels	exclu	ideo
-----	-------------	-------	---------	------------	---------	--------	-------	------

Window Size	Т	T+G	T+B	T+G+B	Count
0.2s	62.4	70.06	99.88	99.91	6713
0.5s	66.04	74.77	99.88	99.88	5990
1s	69.87	77.96	99.9	99.86	4999
2s	54.59	64.35	98.08	98.22	676

Table 3.9: Results from Classifying Participant using 20 fold CV on Touch Frequency (TF); Touch Frequency+Touch pressure-location (TF+T); Touch Frequency+Touch pressure-location+Gaze Frequency (TF+T+GF) feature sets.

Window	TF	TF+T	TF+T+GF
0.2s	53.76	86.8	87.21
0.5s	71.86	90.15	90.48
1.0s	64.25	86.36	87.24
2.0s	87.68	94.82	95.36
(\mathbf{L}) W(4) \mathbf{Q}	<u>a</u> 1 .	• ,	· ,• •

(a) With no gap between instances, emotions in

(b) With 2s Gap between instances, emotions in

Window	TF	TF+T	TF+T+GF					
0.2s	42.46	75.74	76.41					
0.5s	67.35	84.86	84.88					
1.0s	57.21	79.4	80.99					
2.0s	66.87	84.8	85.65					
$(\cdot) \mathbf{W}^{\prime} (\cdot) $	1							

(c) With no gap betw	veen instances,	emotions	out
----------------------	-----------------	----------	-----

Window	TF	TF+T	TF+T+GF
0.2s	40.67	76.92	77.97
0.5s	63.37	83.71	84.78
1.0s	50.42	77.78	79.61
2.0s	83.31	91.84	92.24

(d) With 2s Gap between instances, emotions out

Window	TF	TF+T	TF+T+GF
0.2s	31.43	63.45	65.85
0.5s	60.06	76.98	78.86
1.0s	44.1	69.12	71.35
2.0s	53.43	76	78.38

Touch and Gaze, we ran Weka's Attribute Evaluator using 20-fold CV of the Best First search method on the feature set of Touch and Gaze feature set as described in Section 3.4.2. Note that beyond participant information, the location-median is the most commonly selected feature, though when normalized by feature count per parameter, pressure-based features are the most popular, followed by location-based features. Frequency-based gaze locations are next most popular. Figure 3.7 breaks down each parameter and their relative popularity where each cell represents the number of features of a statistical type selected at each iteration. The most often selected features were the 11 calculated medians of touch location which were chosen 100% of the time during 20-fold CV.

3.4.5 Experienced Emotion Trajectory and Interview Results

Participants provided self-reported emotional state data with respect to Russell's 2D affect grid during two neutralization tasks and following two emotion tasks. Paired t-tests showed no significant difference (p > 0.05) between the neutralization tasks nor an order effect between the emotion tasks. We therefore ignore emotion order for subsequent analysis.

In paired t-tests, we found significant differences in self-reports between neutral and emotion tasks for each of *Stressed*, *Depressed* and *Excited* in both arousal and valence (p < 0.05). The *Relaxed* task did not show significance. A plot of emotion trajectory shows participants' starting state and the movement across the 2D affect grid (Figure 3.8).

Both high arousal emotions (*Excited*, *Stressed*) were consistent with expectations where participants reported a shift in emotion toward the corner of the grid represented by the target emotion word. For some participants, the immediacy or recency of the recalled events really helped to highlight these emotions. As this experiment was run towards end of term, this coincided with final exams and holiday reunions, both cited as reasons for ease of recall.

"I'm leaving to see my family for the first time in three years, I can't stop being excited." – P8

"Excited was easy – the situation was more recent and was more important [than my Depressed memory]." – P22

		Selection %	Median	Min	AUC	Mean	Max	Total Var	Var	Index	Peaks		
	a	100											
	Pressure	95-75											
lch		70-50										(a di	
Tot		45-25											
ssic	_	100											
Cla:	tior	95-75			-								
	oca	70-50											
		45-25											
	Pressure	100											
		95-75				and a second							
-		70-50											
rec		45-25					0						
h.	Location	100											
onc		95-75				1							
Г		70-50											
		45-25											
Gaze Freq	Location	100											
		95-75										a second	
		70-50										l parte	
		45-25											
												Popula weight by feat count	rity ed ure

Figure 3.7: Feature selection popularity by statistic. Pressure-based touch features are most popular, followed by location-based touch and gaze frequency features on location data also third, when normalized against total number of features available by modality.

"I have a lot of school assignments right now and I kind of toggled between many memories [for Stressed]. It was hard to pick one to feel but I think that might have added to the feeling." – P21 "...[W]hen I was doing *Stressed*, I felt like I wanted to punch something it was so gut-wrenching." – P29

The low arousal emotions, *Relaxed* and *Depressed*, moved as expected in valence but not arousal, which remained overall at its neutral "resting" position. In the case of *Relaxed*, this might be explained by perceived similarity between this emotion task and the 'resting' start condition.

"Relaxed was easy to express because it's pleasant and I want to feel it and also, I'm sitting on a couch which helps." – P28, corroborated by P2, P18

For these two emotions, some participants reported that the emotion *Depressed* was linked to *Stressed* in their memories (e.g., , feeling stress about exams was also depressing), which may explain some of the unexpected movement in arousal for *Depressed*. Four participants also reported feelings so strong that their *Depressed* memory evoked active tears, while others indicated that these feelings were somewhat mitigated by the experience of stroking a soft body.

"My [*Depressed*] memory was very clear and I was able to recall a lot of details. It really helped to be touching a soft thing and felt like it was taking some of my sadness." – P29, corroborated by P15, P23

Another possibility for both of these emotion targets is that participants were simply unable to turn down their arousal state to this degree during the relatively short time of the session.

Each participant self-reported a genuineness rating of how authentically they experienced the target affect in each emotion task. On a scale of 1–10 with 1 being *completely contrived or artificial* and 10 being *completely authentic as in the original experience*, participants rated $\mu = 8.29$ ($\sigma = 1.38$) when *Depressed*; $\mu = 8$ ($\sigma = 1.41$) when *Excited*; $\mu = 7.5$ ($\sigma = 1.68$) when *Stressed*; and $\mu = 7.5$ ($\sigma = 1.51$) when *Relaxed*.



Figure 3.8: Changes in individual's self-report of emotion after Neutralization (start) and Emotion tasks (finish); N=14 for *Stressed & Relaxed* and N=16 for *Depressed & Excited*. Overall, we see a move from the origin to the representative quadrant. *Stressed* and *Excited* show the strongest overall change along both **Arousal** and **Valence** axes. *Relaxed* shows the least change with disconnected points referring to "no change" from neutral state.

3.5 Discussion

In this section, we return to our research questions and address our suppositions and research methodology.

3.5.1 Findings

Classification accuracy improves with increased modality support – Accepted.

The accuracy rates of biometrics in CV suggests that any time biometrics can be artfully employed, they should be. For example, skin conductance sensors may be useful for many touch systems. However, with touch and gaze performing at roughly literature accuracies when participant knowledge exists, suggests that emotion classification systems based off of just these modalities would be feasible for at least some applications. Due to the invasiveness of some biometric sensors, such a system has many possible advantages.

The biometric signal features were used as a check against touch and gaze features. For example, in LOO, biometric features performed just as poorly as touch and gaze features, which suggests a high rate of individual differences in the features we have calculated. While there may be generalizable features available in biophysical signals, we are not certain that we have found them here.

Recognition rates will increase alongside greater participant knowledge – Accepted.

Low LOO results imply low generalizability of an individual's emotional behaviour at least when touch, gaze, and biometrics are chosen as the nonverbal channels of analysis. Thus, any system that plans to perform emotion classification is advised to include all expected users in the training pool. As long as users are included as part of a calibration period, the system does not require explicit user identification as evidenced by the relatively strong performance of unlabelled participants over LOO. In cases where the highest accuracy is needed, the system should obtain a priori participant identification as seen in the performance using participant-labelled data.

Reducing sample density reduces classification accuracy – Partially accepted.

From Figures 3.4 and 3.6, increasing window sizes from 0.2s to 2s improves classification under no-gap conditions. However, when we introduced gaps be-

tween instances – where windows were not adjacent – a dropoff in accuracy occurs at 2s windows. Upon closer examination, we notice that the added gaps for 2s windows also decimated the instance count down more than 90%, from 7435 down to 676 instances (see Appendices 3.6b and 3.9b).

In general, larger, adjacent windows result in marginally higher classification accuracies and can be useful in adjusting parameters to system requirements. Relatively minor reduction in accuracy with adjacent versus non-adjacent instances suggests that the same individual touches similarly in same period of emotion expression. It may be possible to decrease computational load during data collection as continuous capture may not be necessary when using short window instances. However, seeing how instance count may influence classification performance, we recommend balancing window sizes with capture length such that short windows are used with systems that require fast response time.

Classification accuracy will be improved with the addition of frequency-domain features – *Rejected.*

At 54Hz, the frequency domain features did not provide a significant improvement in classification performance on touch data. However, the selection of gaze frequency features (see Figure 3.7) suggests that there may be a benefit for gaze data. We looked into frequency features in part due to others' success using frequencybased features [4]. However, with our low sample rate coupled with short windows, this does not appear to be salient for this classification task. Furthermore, frequency domain features require some extra processing by way of Fourier transformations over the standard pressure-location set prior to model building; therefore, where processing time is a priority, the standard feature set may be preferred.

3.5.2 Experimental Methodology

Our experimental methodology reflected our primary imagined use case: a zoomorphic robot designed for therapy. It had several elements that were not standard, including the emotion elicitation method, the choice of emotions investigated, study framing (the setup itself, with the participant interacting with an unresponsive furry object), and various aspects of the analysis. With results in hand, it is relevant to critique each of these aspects for its ability to produce valid data in general, and insights on our research questions in particular.

Emotion Elicitation We were surprised at participants' intensity of expression during the emotion elicitation tasks; although the technique was validated by literature, we elicited stronger emotional reactions from our participants than we expected. The method was shown to be valuable for an experiment run in a laboratory where people can otherwise find it hard to act as they would in more natural settings. We believe that some variation on recall as a method for eliciting emotion can be employed in future studies on our robot.

Emotion Set Although we have reported both high and low classification accuracy rates, we have some skepticism over whether accuracy rates are a good indicator of a successful emotion model. There is certainly value in an accurate system, but there are some underlying assumptions of a discrete classification model that we question.

Here we assume that participants express a roughly *steady state* emotion, felt across the entire memory recall. However, it is possible that strong emotions may be felt only for an instant before autonomic emotion regulation or coping mechanisms take over [39]. The horizon over which we sample a participant's emotional state, and the assumption of immediacy has direct impact on what kind of decisions we would want an interactive system to implement. Our discrete classification system can identify differences in minute-long interactions, but cannot provide us with an accurate estimation of an emotional inflection point (i.e. transition from one emotion to another). A truly interactive system would need to react to the *change* in an emotional state and adapt over many samples.

Furthermore, when engaging in natural emotional conversation, interactions with pets or friends allow for error correction: if we are mistaken on our first interpretation, further context helps us to reassess quickly and correct our language. Working towards an adaptive model (rather than a prescriptive one) would go further into developing a meaningful relationship over a direct and immediate calland-response instructing interaction [89]. Using touch data in context with gaze and biometric analysis lays the groundwork for extending haptic human-robot in-

teractions from instructional directives to meaningful conversational relationships.

3.5.3 Application Implications and Future Work

We return to our example applications as a way of grounding our findings and give concrete ideas of how they could be deployed for use.

Social robot therapy Out of the three studied modalities, touch is the most central for social robot therapy. Our current findings indicate that as long as the user is previously known to the system, distinguishing between four different emotional states can be done quite robustly. This provides intriguing opportunities for further development of therapeutic robots that could automatically run human-affect recognition and adjust their movement based on it. For example, when the user is touching the robot in a way that communicates stress, the robot could alter its breathing behaviour to attempt to calm down the user and reduce possible anxiety.

While gaze and biometrics did improve classification, their use in practical scenarios is more challenging. For robust detection of gaze, the user must always face the robot at a certain angle or wear a calibrated head-mounted gaze tracker. Similarly, biometrics requires instrumentation before reliable readings of signals such as heart rate and skin conductance can be measured. In contrast, touch interaction with the robot typically consists of momentary touch contact that may too short and infrequent for measuring biometric signals through sensors embedded in the robot. However, we believe these sensory systems can be easily integrated where some considerations are met: careful sensor placement for gaze attention and training data collection sessions.

To be used effectively in therapy, an introduction period would be required where an expert such as a therapist would introduce the robot and guide potential users in providing training data for recognition of emotions via touch. Although this does imply a setup cost for use, potential benefits in environments where real animals cannot be used (such as some hospital environments) compensates for the initial time investment. *Entertainment* As the example with the most controlled environment, embedded sensing systems can be integrated into product development, exploiting many modality combinations. We can imagine handheld video game controllers using skin conductance and pressure data because the touch contact with a controller is likely more continuous than that with a social robot. On the other hand, the spatial range of motion in touch gestures sensed with the smaller controller surface may be more limited; in this case, pressure features may provide richer data than location. In addition, it could be possible to develop virtual reality (VR) headsets with gaze tracking and BVP sensors at the temples. Prototyping would be possible with any existing controller that can be augmented with current commercially available sensors.

Since the emotional states experienced with games can be different from those studied in the current paper, it would be vital to run further studies on employing classification models for emotional game play. Fortunately, gaming applications have existing personalization paradigms that can be leveraged for model building: users already log-on with identifying credentials, providing a priori participant label and in-game tutorial sessions offer an opportunity for collecting and building a user emotion model.

Assisted driving Cars are environments where collection of touch, gaze, and biometric data could create a sweet spot of low intrusiveness or annoyance, privacy (not obvious to an observer in or out of the car) and accuracy. Sensors for touch location, pressure, and skin conductance could be integrated into steering wheels. Seat and seat belt could be utilized for heart rate measurements, while rearview mirror and windshield edges offer natural locations for gaze trackers.

Identification of user would not be an issue in the automotive environment where the number of different people using a particular car is typically low. Personal information consisting of touch, gaze, and biometric data could be stored in keyfobs. However, acquiring training data would likely take time, and the assistive features based on acquired emotional data must be downplayed before enough data is collected for reliable classification. A subtle example of utilizing emotional information of the user could be to switch to relaxing music when the driver is detectably stressed.

3.6 Conclusion

In this study, we presented affect classification results from emotionally influenced touch and gaze behaviours, and checked with biometric data. Across the three modalities, data was collected via: a custom piezoresistive fabric touch sensor embedded in a furry football-sized stationary robot; a Tobii EyeX gaze tracker; and Thought Technology's BioInfiniti Physiology Suite of skin conductance, respiratory rate, and heart rate variability (by way of Blood Volume Pulse) sensors. Our data set is composed of sensory data as well as self-reports of emotion genuineness and intensity as participants recalled intense emotional memories spanning Russell's 2-D arousal-valence affect space, namely Depressed, Excited, Stressed, and *Relaxed*. For models trained with test participant data using pressure-location features, the overall emotion recognition rate was roughly 83%, 87%, and 99% for touch, touch + gaze, and touch + gaze + biometrics respectively. Performance drops steeply when test participants were left out of the training model, resulting in 31%, 31%, and 29%, approaching chance (25%). We tried increasing the feature set by incorporating frequency features for touch and gaze modalities, 79%, 85%, and 85% respectively for touch frequency features, frequency and pressurelocation touch features, and touch frequency, touch pressure-location, and gaze frequency features combined where LOO performs similarly poorly at 30%, 32%, and 35% respectively.

These results inform design of a therapeutic social robot embedded with realtime emotion classification and we make the following recommendations:

Emotional behaviour encoded in touch and gaze interaction is sufficient: While including biometric data greatly improves accuracy, current technology requires they be worn rather than embedded, resulting in a more restrained experience. Setup restrictions may interfere with natural emotional expression and sensors affixed to the hand and body can feel invasive.

A training or calibration phase increases the model's prediction ability: Increasing participant information greatly improves the classification model's prediction accuracy. While this stage likely requires guidance from an expert or therapist, the initial training investment facilitates the learning of user-specific characteristics and develops a more robust user behaviour model, thereby allowing for a personalized and productive experience.

Sampling density and feature count may be reduced to improve computation load: During real-use cases, the speed of classification and reaction is a serious concern. Our findings indicate that interruptions in data collection at up to 2s intervals may be tolerable. Conversely, a recognition system that detects high frequency behaviours could dynamically update sample density.

Although we have achieved possibly usable classification rates, our reflections lead us to believe that a categorical affect model has clear limitations that must be addressed. People do not experience emotions in isolation nor discretely, rather emotional experiences follow a trajectory with distinctive peaks and valleys. Future detection systems need to develop models that follow the rise and resolution of an experience. While this study used a stationary robot, a deployed interactive system must acknowledge that its response has influence over user emotional reaction, necessitating dynamic adjustments to behaviour modelling.

Chapter 4

Behaviour Sketching

Previous chapters have focused on how a robot senses human affect and intent; now we look to close the interaction loop and consider how a machine can convey recognizable emotions to the human user.

Robots that interact directly with people will soon become commonplace [29, 66], from manufacturing [36] to healthcare and the home [14]. Such machines must function with a degree of social intelligence, and for many applications, render and react to affect via touch and physical gesture [28, 34].

Both the Haptic Creature [108] and CuddleBot [3] were created to study emotional touch and its therapeutic benefits. They use exo- or endo-skeletal, vibratory, heat or pneumatic elements, and sophisticated signal processing and control requiring powerful computation and architecture. Their high expressive potential (via, for example, breathing, purring, hunching, and head movement) requires complicated coordination of single element motion. Inspired by research on emotional breathing [13, 88], we zeroed in on 1-DOF breathing behaviours in two distinct robot form factors to discern what factors in motion are emotionally suggestive. Originally organized as a case study of periodic breathing, we sketch and have users (N=20) evaluate behaviours on palm-sized, limbless 1-DOF robots collectively dubbed CuddleBits—flexible, furry, and fully-covered FlexiBit, and the rigid, wooden, and exposed RibBit (see Figure 4.1).



Figure 4.1: The rigid RibBit (Left) and fur-covered FlexiBit (Right) explore very different form factors using similar actuation principles and requirements. Both can be compressed without damage, allowing for a more naturalistic haptic display.

4.1 Related Work

While physical form is suggestive of emotional traits, we borrow from other animation methods to suggest anthropomorphism and increase expressivity of inanimate objects.

Animation of emotion: Attribution theory [56] suggests that humans find agency in many objects and motions, supporting the communicative viability of very simple forms. Conveying emotion on non-humanoid forms has been a mainstay of visual animation from its beginnings, illustrated with Disney's 'sack of flour' exercise (http://tinyurl.com/pjhwrhg) where artists breathe life into a humble bag [99]. Believable emotion display does not require realistic rendering of the animal or inanimate object conveying it, only a recognisable anthropomorphic movement [99, 106].

Emotions and robot believability: To be believable social agents, robots should seem capable of emotional processing and expression [11]. Many such robots have

been built in zoomorphic form with encouraging results [67, 86, 101, 107]; but these forms are much more complex than single degrees-of-freedom.

Animation concepts have been successfully applied to *physical* expression or human inference of affective parameters on non-realistic everyday objects, e.g. the ambient influence of a stick-like sculpture's movement on a desk-workers activity [47], and a physically animated phone's portrayal of emotions spanning Russell's 2-dimensional affective grid [30, 83] or an expected liveliness [75]. Expressive animations produced primarily for touching are less common.

Physiological and emotive impact of breathing: Direct physical contact with another's breathing motion affects physiology; e.g., in skin-to-skin contact therapy for premature infants it promotes physiological stabilization [65]. Similar effects are seen with touch-based social robots [103]. A robot's felt respiratory motion can reliably impart a physiologically and subjectively significant calming influence [88].

Human breathing behaviours reflect affective state [13], and breathing is an expressive visual animation tool able to capture drowsiness to distress. The Haptic Creature's breathing display was crucial to being able to convey emotion [109]. The present work tests the ability of breath-like motion alone to represent a full emotional range.

4.2 Experiment Method

We recruited 20 participants (11 male, 8 female, 1 other), aged 20–40 with cultural backgrounds from North America, Europe, Southeast Asia, Middle East and Africa. All participants had completed at least an undergraduate degree and were compensated \$5 for the 30 minute study.

Participants were seated and invited to inspect the inanimate robots, then instructed to use one hand to touch the robot and the other to use the mouse (Figure 4.2). They were given the task of rating each behaviour on a 5-point semantic differential (-2 Mismatch to +2 Match) for four situations where the robot was stressed, excited, depressed, or relaxed (see Figure 4.3). For instance, for "FlexiBit feels stressed", a participant would play the behaviour and rate how well it matched the robot portraying stress.



Figure 4.2: Experimental setup showing a participant touching the FlexiBit and rating behaviours. The screen's quadrants present the four situation descriptions.

During playback and rating, hands obscured participants' view of the robot; motion was experienced largely haptically. Noise-cancelling headphones played pink noise to mask mechanical noises; instructions were communicated by microphone.

Ratings for each robot were performed separately. Robot block order was counterbalanced, with a 2m rest. For each block, all four emotions were presented on the same screen so participants could compare globally. Behaviours (15s clips) could be played at will during the block.

Order of behaviours and emotion was randomised by participant for the first robot. To reduce cognitive load, participants saw the same behaviour/emotion order for the second block. In total, each participant performed 64 ratings (8 behaviours \times 4 emotions \times 2 robots). Each session took \sim 30m including a post-experiment



Figure 4.3: Close-up of the interface participants used to rate behaviours.

interview.

Quantitative data included situation ratings and completion time, estimated by duration of mouse focus within quadrant. In addition to the 64 behaviour ratings per participant, we also recorded the time it took for each participant to complete ratings per situation. This was estimated by adding up the amount of time that the mouse cursor was in each of the four quadrants of the interface. We approximated task difficulty with time spent evaluating behaviours, suggesting challenge in aligning robot behaviours with emotion.

4.3 Results

We ran pairwise Wilcoxon signed-rank tests with Bonferroni correction. Ratings of the two designed behaviours for the same situation showed no significant differences ($\alpha = .050/8 = .006$; all p's $\ge .059$). Thus, we averaged ratings into four pairs by emotion target (e.g., (1) & (2) in Figure 4.4); pairs appear on y-axis of 4.5). The x-axis displays the four emotions. Darker colours indicate higher participant ratings, and in an ideal case (where participants think the behaviours match the situations researchers designed them for) the darkest colours appear on the diagonal.



Figure 4.4: Waveforms of behaviours as designed by researchers.

The behaviour ratings are grouped based on the situation the behaviour was designed for and the situation for which the behaviours were rated. The ratings are grouped by the intended representative emotion and the emotional content for which the behaviours were rated. Darker colours on the diagonal indicate that where behaviour ratings matched the design intention the behaviours matched the situations that the behaviours were designed for. For example, it can be seen that the designed behaviours for *Stressed* were rated to be a better match for the excited situation.*Excited*).

Effect of situation on behaviour ratings. Friedman's test on behaviour ratings showed significant differences between behaviours per situation for both robots (all p's < .001). Post hoc analyses using Wilcoxon signed-rank tests were conducted with a Bonferroni correction ($\alpha = .050/6 = .008$) to further analyse the effect of **situation condition** on *researcher-designed behaviours* (Figure 4.6):

– Stressed, Excited, or Relaxed: Significant differences between high and low arousal behaviours (*Stressed-Depressed*, *Stressed-Relaxed*, *Excited-Depressed* and *Excited-Relaxed*, all p's \leq .002). No significant differences between behaviours with the same arousal level but different valence content.

- **Depressed**: No significant differences between three high and low arousal behaviour pairs. A significant difference between behaviours with the same arousal level but different valence content (*Stressed-Excited*, $p \le .007$).

For three of the four situation conditions, participant ratings of behaviours



Figure 4.5: Mean behaviour ratings for FlexiBit grouped by the researcherdesigned behaviours (horizontal) and the situation for which the behaviours were rated by participants (vertical). Researcher-designed behaviours correspond with (a) to (h) in Fig. 4.4.

were decisive (high color contrast between behaviour ratings on Figure 4.5 y-axis). Specifically, by **situation condition** and *researcher-designed behaviours*:

For situation = **Relaxed**, **Excited**, or **Stressed**, pairwise comparisons between all researcher-generated low and high arousal behaviours showed significant differences in ratings (*Depressed-Excited*, *Depressed-Stressed*, *Relaxed-Excited* and *Relaxed-Stressed*, all p's \leq .002). No significant differences were found for ratings of behaviours with the same arousal level but different valence content: *Depressed-Relaxed* and *Stressed-Excited* (p \geq .017).

For situation = **Depressed**, pairwise comparisons showed significant differences between the following low and high arousal behaviour pairs with both robots: *Depressed-Excited* and *Relaxed-Excited* (p's \leq .001). Also, a significant difference

		Flex	(iBit		RibBit			
	S	Е	D	R	S	E	D	R
S-E	.391	.142	.159	.076	.759	.037	.007	.017
S-D	.001	.000	.004	.000	.000	.000	.012	.000
S-R	.001	.000	.014	.001	.001	.002	.032	.001
E-D	.000	.000	.000	.000	.000	.000	.000	.000
E-R	.000	.000	.000	.000	.000	.000	.000	.000
D-R	1.000	.713	.668	.501	.582	.270	.713	.668

Figure 4.6: Pairwise comparison p-values (Wilcoxon) of behaviours (row) for different situation conditions (col), sig. diff. are darker. Notice RibBit(S-E, D): in the **Depressed** condition, *Stressed* and *Excited* were rated significantly differently.

was found between the low and high arousal pair *Depressed-Stressed* with FlexiBit ($p \le .004$). With RibBit, a significant difference was found between the positive and negative valence pair *Excited-Stressed* ($p \le .007$).

This held for **Depressed** with some exceptions: no significant differences were found with either robot when comparing ratings of *Relaxed* and *Stressed* behaviours (p's \ge .014). In addition, with RibBit, no significant differences were found for the ratings of *Depressed* and *Stressed* behaviours (p = .012), however, comparison of *Excited* and *Stressed* revealed a significant difference (p \le .007).

Effect of robot on behaviour ratings (not significant). Wilcoxon signed-rank tests using Bonferroni correction showed no significant differences in ratings between the two robots ($\alpha = .050/16 = .003$; all p's $\ge .026$).

Duration (not significant). A two-way (2 robots \times 4 situations) repeated measures ANOVA showed no significant difference in the time spend on rating behaviours.

4.4 Discussion

We address our findings with respect to our hypotheses.

Hypothesis 1: Different levels of arousal are easier to interpret than different levels

of valence. -Accepted

In general, participants were able to perceive differences in behaviours designed to convey high or low arousal. The main parameter for communicating arousal variations and most commonly recognized by participants in our behaviour design was frequency. Speed or frequency was most mentioned as having communicated arousal variation: low arousal from low frequency and high arousal from high frequency. This confirms that this 1-DOF display is able to reproduce earlier findings [30, 80, 110]. High or low physical activation signals are easily distinguishable and are good indicators of alertness, evidenced by results where consistent arousal states were well-matched.

As hypothesised, participants were less able to interpret valence from robot movement. This has also been a challenge for other physical displays [30, 80]. Possible reasons include: ineffective behavioural language for valence polarity (non-periodic, asymmetric signal shapes); breathing as a behaviour might not naturally convey valence variations and/or additional DFs are needed to disambiguate them; materiality played a role (less likely considering consistency between our two prototypes).

Unexpectedly, ratings for "depressed" situation diverged significantly.

Interviews suggest two reasons: (a) **Depressed** was being conflated with **Stressed**. Participants reported experiencing both emotions in concert or as a result of the other. And, (b) breathing (by RibBit in particular) did not have the ability to express depression for some participants. Others simply were not convinced that the robots, and the RibBit in particular, could express a depressed state via breathing behaviour alone.

Suggestions to improve believability and differentiability for **Stressed** included sighing and avoidance actions like retreating or turning away. Out of scope for this paper, this will inform future behavioural and actuation design.

Hypothesis 2: FlexiBit's behaviour will be perceived as conveying more positive valence than RibBit's due to its softer and more pleasant feel. *—Rejected*

Post-study interviews revealed that participants reported the movement expressed by the two robot forms as sensorially though not necessarily emotionally different. FlexiBit felt nicer to touch, but the motion was less precise than that of the RibBit. RibBit's movements were interpreted as breathing or a heartbeat despite the exposed inner workings reducing the 'lifelikeness' of the forms.

Unexpectedly, while participants specified preferences for FlexiBit's fur and RibBit's motor precision, pairwise comparisons of the same emotions revealed no significant difference between robots. Movement rather than materiality dominated how participants interpreted emotional expression; although visual access to form was restricted during movement, tactility might have modulated perception of, e.g., life-likeness.

Chapter 5

Conclusions and Future Work

This thesis explores one full iteration of the HRI loop in affective touch communication using a custom-built sensor for our CuddleBot family of therapeutic robot pets through three distinct studies, each described in its own chapter. We tested our sensor on classification of gestural touch ($N_1 = 10$, $N_2 = 16$) with results at 79% - 95% accuracy (chance 14%) depending on noise factors, consistent with literature values. Using the same custom touch sensor, we ran affect classification on multimodal experienced-emotion data (N = 30) with overall accuracy rates of 83%, 87%, and 99% accuracy (chance 25%) on touch-only, touch + gaze, and touch + gaze + biometric data respectively where random forest trained models included test participants. And finally, as breathing characteristics have been shown to have recognizable emotional properties [13], we asked participants (N = 20) to label and rate emotional breathing behaviours on two distinct 1-DOF robots. Results showed that high arousal designs (Stressed and Excited) were significantly recognizable over low arousal designs (Depressed and Relaxed); distinctions from negative valence (Stressed and Depressed) to positive valence (Excited and Relaxed) were more difficult.

From automatic affect detection to human-recognizable robot affective robot behaviours, we have demonstrated the feasibility of the full HRI loop. The following outlines the impacts of each study and describes future work that builds on findings from this thesis.

5.1 Outcomes and Impacts

Each of the three aforementioned studies forms a chapter of this thesis and is either a previously published work (Chapter 2 [19]), in preparation for publication (Chapter 3), or part of a larger ongoing study (Chapter 4). In each case, the impacts extend beyond this thesis – from informing research directions within the lab to fueling interests of the larger community.

Gesture Recognition and Touch Sensor:

The biggest impact arising from our conference paper [19] reported in Chapter 2 is that of the ICMI Grand Challenge. Led by Merel Jung of the University of Twente in the Netherlands, we published two datasets, one from our work dubbed Human-Animal Affective Robot Touch (HAART), the other previously published and titled the Corpus of Social Touch (CoST) [52]. The challenge posed to the community involved developing a classification model to better understand or improve on dataset authors' work. We received 10 submissions from around the world with subject-independent ¹results ranging from 35% to 71% accuracy rate using Support Vector Machines (SVM), neural nets, and random forest, to name a few (results lower than the subject-dependent values published by us). Of these, four teams presented papers discussing their techniques and results at ICMI '15 in Seattle WA [53].

More locally, we are planning both improvements and full redesigns of our custom sensing apparatus. While the fabric sensor has served us well for data collection, we can't help but explore questions on biophysical requirements of polling rate and resolution. While human control of movement is roughly 5-10Hz [92], human recognition of tactile sensation is degrees of magnitude higher at up to 10kHz [72]. So while a robot skin that polls at 54Hz will capture human movement, it may yet fall short of human sensory ability. On the other hand, this level of sensing would overwhelm computational load and it's not clear whether rates this high are even necessary. Current planned studies are aimed at determining the

¹Subject-independence indicates that classification models know nothing of test participants as models are trained and tested on data sets with mutually exclusive sets of participants e.g., P1's data appears in either the training set OR the test set but not both. In contrast, subject-dependent testing refers to training and testing on the same participant where the model may learn something of participant behaviour.
required range of resolution and polling rate of tactile sensing skins. By exploring sensing mechanisms, our lab has generated a number of new tactile sensors using conductive paint, weaving with conductive thread and yarn, as well as silicone-embedded capacitors (developed in collaboration with UBC engineering students).

Affect Recognition:

A version of Chapter 3's data collection and work on classification of emotion is targeted for an upcoming journal deadline. In the longer term, it has also inspired further studies for determining emotional trajectory. We considered steady state emotions to begin our exploration of how to classify affective states, however, this is only a simplified model of the full emotional experience [17]. Emotion regulation is an important vehicle for coping with negative feelings, a natural but highly individualistic vehicle that will influence the emotional path [40]. We plan an investigation into the artifacts of touch that may help us understand if and when a change occurs, which may allow us to develop robot systems that act as a catalyst to hasten or improve that trajectory to more positive-valence emotions.

Affective Robot Behaviour:

The study of haptically recognizable affective robot behaviours leads us to analyze breathing patterns [13], heavily due to it's calming effects [88]. The findings in Chapter 4 are part of a larger series of studies that further explore the range of affective breathing in by expanding the emotion set both in design and interpretation. These behaviours will form a more complete set of complex affective robot reactions to human input. In the meantime, the sketching and generation of these behaviours have been demonstrated at two venues already: once as a mapping from gesture to reaction as a demo at ICMI'15 in Seattle USA [18] and then as a behaviour display vehicle at EuroHaptics'16 in London UK [15].

5.2 Future Work

The robots used in the studies of this thesis range from simple 1-DOF haptic displays of the CuddleBits to the lap-sized 5-DOF modular form of the CuddleBot. Breathing behaviours created on the CuddleBits not only help to form our impressions of affective reaction recognition, but also to inform further refinement on the larger, more mature CuddleBot's requirements. Beyond breathing mechanisms, we are also exploring many of the CuddleBot's degrees of freedom independently; as posture is also highly expressive [26], we target spinal movements that effectively create curling and stretching behaviours to reflect fear or dismay and relaxation or welcome respectively.

Improving affect detection and building a complex set of believable and recognizable affective behaviours begins our expansion of Yohanan's HRI loop [110]. The original interaction loop suggests a naive model wherein human output is assessed and categorized, mapping to a robot reaction. However, this is not how we expect to interact with each other nor how to communicate emotionally with our animals. The most natural affective exchanges follow a more conversational model [16, 89] which makes use of error correction as well as posterior maximum likelihood calculations to develop smarter behaviour iterations that acknowledge human adjustments to displayed behaviours. Future work towards real-time use of our therapy robot will include the creation of a more robust HRI model detailing a probabilistic decision process to determine the most appropriate robot reaction to human behaviour.

Bibliography

- R. Adolphs. Neural systems for recognizing emotion. *Current opinion in neurobiology*, 12(2):169–177, 2002. → pages 38
- [2] J. Allen and K. E. MacLean. Personal space invaders: Exploring robot-initated touch-based gestures for collaborative robotics. In *Proc* ACM/IEEE Int'l Conf on Human-Robot Interaction - Extended Abstracts, pages 185–186, 2015. → pages 18
- [3] J. Allen, L. Cang, M. Phan-Ba, A. Strang, and K. MacLean. Introducing the cuddlebot: A robot that responds to touch gestures. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 295–295. ACM, 2015. → pages 5, 7, 13, 81
- [4] K. Altun and K. E. MacLean. Recognizing affect in human touch of a robot. *Pattern Recognition Letters*, 66:31–40, 2015. → pages 2, 8, 10, 36, 39, 44, 45, 48, 55, 57, 58, 64, 75
- [5] B. M. Appelhans and L. J. Luecken. Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229, 2006. → pages 46
- [6] B. M. Appelhans and L. J. Luecken. Heart rate variability and pain: Associations of two interrelated homeostatic processes. *Biological Psychology*, 77(2):174–182, 2008. ISSN 03010511. doi:10.1016/j.biopsycho.2007.10.004. → pages 46, 65
- [7] B. D. Argall and A. G. Billard. A survey of tactile human–robot interactions. *Robotics and Autonomous Systems*, 58(10):1159–1176, 2010.
 → pages 18, 19
- [8] M. R. Banks and W. A. Banks. The effects of animal-assisted therapy on loneliness in an elderly population in long-term care facilities. *The journals*

of gerontology series A: biological sciences and medical sciences, 57(7): M428–M432, 2002. \rightarrow pages 39

- [9] M. R. Banks, L. M. Willoughby, and W. A. Banks. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. J American Medical Directors Assoc, 9(3):173–177, 2008. → pages 18
- [10] S. B. Barker and K. S. Dawson. The effects of animal-assisted therapy on anxiety ratings of hospitalized psychiatric patients. *Psychiatric services*, 1998. → pages 1, 39
- [11] J. Bates et al. The role of emotion in believable agents. Communications of the ACM, 37(7):122–125, 1994. → pages 82
- M. Bekoff and J. Goodall. *The emotional lives of animals: A leading scientist explores animal joy, sorrow, and empathy–and why they matter.* New world library, 2007. → pages 2
- [13] S. Bloch, M. Lemeignan, and N. Aguilera-T. Specific respiratory patterns distinguish among human basic emotions. *Intl J Psychophysiology*, 11(2): 141–154, 1991. → pages 9, 81, 83, 91, 93
- [14] J. Broekens, M. Heerink, and H. Rosendal. Assistive social robots in elderly care: a review. *Gerontechnology*, 8(2):94–103, 2009. → pages 81
- [15] P. Bucci, L. Cang, M. Chun, D. Marino, O. Schneider, H. Seifi, and K. E. MacLean. Cuddlebits: an iterative prototyping platform for complex haptic display. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 2016. → pages 7, 93
- [16] B. R. Burleson. The experience and effects of emotional support: What the study of cultural and gender differences can tell us about close relationships, emotion, and interpersonal communication. *Personal Relationships*, 10(1):1–23, 2003. → pages 94
- [17] R. A. Calvo, S. D'Mello, J. Gratch, and A. Kappas. *The Oxford handbook of affective computing*. Oxford University Press, USA, 2014. → pages 93
- [18] L. Cang, P. Bucci, and K. E. MacLean. Cuddlebits: Friendly, low-cost furballs that respond to touch. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 365–366. ACM, 2015. → pages 7, 93

- [19] X. L. Cang, P. Bucci, A. Strang, J. Allen, K. MacLean, and H. Liu. 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*, pages 147–154. ACM, 2015. → pages viii, 5, 13, 41, 48, 53, 54, 55, 57, 58, 61, 92
- [20] J. Chang, K. MacLean, and S. Yohanan. Gesture recognition in the haptic creature. In *Eurohaptics*, pages 385–391, 2010. → pages 1, 19
- [21] W.-L. Chang, S. Šabanovic, and L. Huber. Use of seal-like robot paro in sensory group therapy for older adults with dementia. In *Proceedings of the* 8th ACM/IEEE international conference on Human-robot interaction, pages 101–102. IEEE Press, 2013. → pages 6
- [22] J. A. Coan and J. J. Allen. Handbook of emotion elicitation and assessment. Oxford university press, 2007. → pages 44, 51
- [23] J. Cohen. A power primer. Psychological bulletin, 112(1):155, 1992. \rightarrow pages 27
- [24] K. M. Cole, A. Gawlinski, N. Steers, and J. Kotlerman. Animal-assisted therapy in patients hospitalized with heart failure. *American J of Critical Care*, 16(6):575–585, 2007. → pages 1, 3, 18
- [25] C. Conati. Intelligent tutoring systems: New challenges and directions. In IJCAI, volume 9, pages 2–7, 2009. → pages ii, 8
- [26] M. Coulson. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of nonverbal behavior*, 28(2):117–139, 2004. → pages 94
- [27] J. R. Crawford and J. D. Henry. The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 43(3): 245–265, 2004. → pages 43
- [28] K. Dautenhahn. Design spaces and niche spaces of believable social robots. In Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on, pages 192–197. IEEE, 2002. → pages 81
- [29] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Trans of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007. → pages 81

- [30] J. Q. Dawson, O. S. Schneider, J. Ferstay, D. Toker, J. Link, S. Haddad, and K. MacLean. It's alive!: exploring the design space of a gesturing phone. In *Proceedings of Graphics Interface 2013*, pages 205–212. Canadian Information Processing Society, 2013. → pages 83, 89
- [31] L. L. Di Stasi, D. Contreras, J. J. Canas, A. Cándido, A. Maldonado, and A. Catena. The consequences of unexpected emotional sounds on driving behaviour in risky situations. *Safety science*, 48(10):1463–1468, 2010. → pages 1
- [32] P. Ekman, R. W. Levenson, and W. V. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210, 1983. → pages 8, 36, 45
- [33] A. Flagg and K. MacLean. Affective touch gesture recognition for a furry zoomorphic machine. In *Proc Intl Conf on Tangible, Embedded and Embodied Interaction*, pages 25–32, 2013. → pages 1, 8, 10, 13, 18, 19, 20, 21, 23, 25, 27, 40, 41, 45, 48, 55, 57, 58, 61
- [34] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166, 2003. → pages 38, 81
- [35] G. J. Gelderblom, R. Bemelmans, N. Spierts, P. Jonker, and L. De Witte. Development of paro interventions for dementia patients in dutch psycho-geriatric care. In *International Conference on Social Robotics*, pages 253–258. Springer, 2010. → pages 6
- [36] B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar. Gestures for industry: intuitive human-robot communication from human observation. In *Proc. ACM/IEEE Int'l Conf on Human-robot interaction*, pages 349–356, 2013. → pages 1, 18, 81
- [37] P. Good. Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer Science & Business Media, 2013. → pages 25
- [38] K. Goris, J. Saldien, I. Vanderniepen, and D. Lefeber. The huggable robot probo, a multi-disciplinary research platform. In *International Conference* on Research and Education in Robotics, pages 29–41. Springer, 2008. → pages 6
- [39] J. J. Gross. The emerging field of emotion regulation: an integrative review. *Review of general psychology*, 2(3):271, 1998. → pages 76

- [40] J. J. Gross. Handbook of emotion regulation. Guilford publications, 2013. → pages 93
- [41] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. ACM SIGKDD Explorations Newsletter, 11(1):10–18, 2009. → pages 25, 57
- [42] J. T. Hancock, K. Gee, K. Ciaccio, and J. M.-H. Lin. I'm sad you're sad: emotional contagion in cmc. In *Proceedings of the 2008 ACM conference* on Computer supported cooperative work, pages 295–298. ACM, 2008. → pages 43
- [43] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, and A. R. Jaskolka. Touch communicates distinct emotions. *Emotion*, 6(3):528, 2006. \rightarrow pages 1, 18, 36, 44, 45, 48
- [44] M. J. Hertenstein, R. Holmes, M. McCullough, and D. Keltner. The communication of emotion via touch. *Emotion*, 9(4):566, 2009. \rightarrow pages 18
- [45] R. Hill. Perceptual attention in virtual humans: Toward realistic and believable gaze behaviors. In *Proceedings of the AAAI Fall Symposium on Simulating Human Agents*, pages 46–52, 2000. → pages 46
- [46] K. Inoue, K. Wada, and R. Uehara. How effective is robot therapy?: Paro and people with dementia. In *Intl Fed for Medical and Biological Engineering*, pages 784–787, 2012. → pages 1, 5, 6, 7, 10, 18
- [47] N. Jafarinaimi, J. Forlizzi, A. Hurst, and J. Zimmerman. Breakaway: an ambient display designed to change human behavior. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1945–1948. ACM, 2005. → pages 83
- [48] N. Jaques, C. Conati, J. M. Harley, and R. Azevedo. Predicting affect from gaze data during interaction with an intelligent tutoring system. In *Intelligent Tutoring Systems*, pages 29–38. Springer, 2014. → pages 1, 8, 36, 39, 41, 46
- [49] C. M. Jones and T. Troen. Biometric valence and arousal recognition. In Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces, pages 191–194. ACM, 2007. → pages 46

- [50] H. Y. Joung and E. Y.-L. Do. Tactile hand gesture recognition through haptic feedback for affective online communication. In *Universal Access in Human-Computer Interaction*, pages 555–563. Springer, 2011. → pages 15, 17, 18, 34
- [51] M. M. Jung. Towards social touch intelligence: developing a robust system for automatic touch recognition. In *Proc Intl Conf on Multimodal Interaction*, pages 344–348, 2014. → pages 1, 19, 33, 40, 45, 58
- [52] M. M. Jung, R. Poppe, M. Poel, and D. K. Heylen. Touching the void–introducing cost: Corpus of social touch. In *Proc Intl Conf on Multimodal Interaction*, pages 120–127, 2014. → pages 18, 20, 27, 92
- [53] M. M. Jung, X. L. Cang, M. Poel, and K. E. MacLean. Touch challenge'15: Recognizing social touch gestures. In *Proceedings of the* 2015 ACM on International Conference on Multimodal Interaction, pages 387–390. ACM, 2015. → pages 92
- [54] M. Kanamori, M. Suzuki, and M. Tanaka. Maintenance and improvement of quality of life among elderly patients using a pet-type robot. *Japanese J. Geriatrics*, 39(2):214–218, 2002. → pages 1
- [55] J. Kangas, J. Rantala, D. Akkil, P. Isokoski, P. Majaranta, and R. Raisamo. Both fingers and head are acceptable in sensing tactile feedback of gaze gestures. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pages 99–108. Springer, 2016. → pages 49
- [56] H. H. Kelley. Attribution theory in social psychology. In *Nebraska symposium on motivation*. University of Nebraska Press, 1967. → pages 82
- [57] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(12):2067–2083, 2008. → pages 41, 44, 46, 47, 61, 64, 65
- [58] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995. → pages 26
- [59] S. G. Koolagudi, N. Kumar, and K. S. Rao. Speech emotion recognition using segmental level prosodic analysis. In *Devices and Communications* (*ICDeCom*), 2011 International Conference on, pages 1–5. IEEE, 2011. → pages 36

- [60] J. Kortelainen, S. Tiinanen, X. Huang, X. Li, S. Laukka, M. Pietikainen, and T. Seppanen. Multimodal emotion recognition by combining physiological signals and facial expressions: a preliminary study. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5238–5241. IEEE, 2012. → pages 36, 46
- [61] A. Kurauchi, W. Feng, A. Joshi, C. Morimoto, and M. Betke. EyeSwipe. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, pages 1952–1956, New York, New York, USA, 2016. ACM Press. ISBN 9781450333627. doi:10.1145/2858036.2858335. URL http://doi.acm.org/10.1145/2858036.2858335http: //dl.acm.org/citation.cfm?doid=2858036.2858335. → pages 49
- [62] R. W. Levenson. Autonomic nervous system differences among emotions. Psychological science, 3(1):23–27, 1992. → pages 8, 36, 45
- [63] J. Marescaux, J. Leroy, M. Gagner, F. Rubino, D. Mutter, M. Vix, S. E. Butner, and M. K. Smith. Transatlantic robot-assisted telesurgery. *Nature*, 413(6854):379–380, 2001. → pages 7
- [64] P. Marti, A. Pollini, A. Rullo, and T. Shibata. Engaging with artificial pets. In Proc. Annual Conf European Assoc Cognitive Ergonomics, pages 99–106, 2005. → pages 18
- [65] A. J. Mitchell, C. Yates, K. Williams, and R. W. Hall. Effects of daily kangaroo care on cardiorespiratory parameters in preterm infants. *Journal* of neonatal-perinatal medicine, 6(3):243–249, 2013. → pages 83
- [66] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft. Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In *Proceedings of the* 2014 ACM/IEEE international conference on Human-robot interaction, pages 334–341. ACM, 2014. → pages 81
- [67] M. C. Moy. Gesture-based interaction with a pet robot. In AAAI/IAAI, pages 628–633, 1999. → pages 83
- [68] K. Nakajima, Y. Itoh, Y. Hayashi, K. Ikeda, K. Fujita, and T. Onoye. Emoballoon. In Advances in Computer Entertainment, pages 182–197. Springer, 2013. → pages 27

- [69] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6(4):385–394, 2015. doi:10.1109/TAFFC.2015.2432810. → pages 46
- [70] C. Nguyen and F. Liu. Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems CHI '16*, pages 2093–2097, New York, New York, USA, 2016. ACM Press. ISBN 9781450333627. doi:10.1145/2858036.2858137. URL http://doi.acm.org/10.1145/2858036.2858137.http://dl.acm.org/citation.cfm?doid=2858036.2858137. → pages 49
- [71] J. Odendaal. Animal-assisted therapymagic or medicine? Journal of psychosomatic research, 49(4):275–280, 2000. → pages 39
- [72] M. A. Otaduy and M. C. Lin. High fidelity haptic rendering. Synthesis Lectures on Computer Graphics and Animation, 1(1):1–112, 2006. → pages 58, 92
- [73] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1–135, 2008. → pages 36
- [74] T. Partala and V. Surakka. Pupil size variation as an indication of affective processing. *Int. J. Hum.-Comput. Stud.*, 59(1-2):185–198, July 2003. ISSN 1071-5819. doi:10.1016/S1071-5819(03)00017-X. URL http://dx.doi.org/10.1016/S1071-5819(03)00017-X. → pages 36, 46
- [75] E. W. Pedersen, S. Subramanian, and K. Hornbæk. Is my phone alive?: a large-scale study of shape change in handheld devices using videos. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2579–2588. ACM, 2014. → pages 83
- [76] E. Peper, R. Harvey, I.-m. Lin, H. Tylova, and D. Moss. Is There More to Blood Volume Pulse Than Heart Rate Variability, Respiratory Sinus Arrhythmia, and Cardiorespiratory Synchrony ? *Biofedback*, 35(2):54–61, 2007. ISSN 10815937. → pages 46
- [77] H. Perner-Wilson and M. Satomi. Diy wearable technology. In ISEA 15th International Symposium on Electronic Art, 2009. → pages 13
- [78] D. Proulx. Animal-assisted therapy. Critical care nurse, 18(2):80, 1998. \rightarrow pages 1

- [79] P. Rani, C. Liu, N. Sarkar, and E. Vanman. An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9(1):58–69, 2006. → pages 18
- [80] J. Rantala, K. Salminen, R. Raisamo, and V. Surakka. Touch gestures in communicating emotional intention via vibrotactile stimulation. *Intl J Human-Computer Studies*, 71(6):679–690, 2013. → pages 89
- [81] N. E. Richeson. Effects of animal-assisted therapy on agitated behaviors and social interactions of older adults with dementia. *American journal of Alzheimer's disease and other dementias*, 18(6):353–358, 2003. → pages 39
- [82] B. Robins, K. Dautenhahn, F. Amirabdollahian, F. Mastrogiovanni, and G. Cannata. Developing skin-based technologies for interactive robots-challenges in design, development and the possible integration in therapeutic environments. In *IEEE ROMAN*, pages xxii–xxiii, 2011. → pages 19
- [83] J. A. Russell, A. Weiss, and G. A. Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989. → pages ii, 43, 51, 54, 83
- [84] J. Sabourin, B. Mott, and J. C. Lester. Modeling learner affect with theoretically grounded dynamic bayesian networks. In *Affective computing* and intelligent interaction, pages 286–295. Springer, 2011. → pages 44
- [85] J. Saldien, K. Goris, S. Yilmazyildiz, W. Verhelst, and D. Lefeber. On the design of the huggable robot probo. J. Physical Agents, 2(2):3–12, 2008. → pages 18
- [86] J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaeillie, and D. Lefeber. Expressing emotions with the social robot probo. *Intl J of Social Robotics*, 2(4):377–389, 2010. → pages 5, 6, 10, 83
- [87] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*, pages 71–78, New York, New York, USA, 2000. ACM Press. ISBN 1581132808. doi:10.1145/355017.355028. URL http://portal.acm.org/citation.cfm?doid=355017.355028. → pages 56
- [88] Y. Sefidgar, K. E. MacLean, S. Yohanan, M. Van der Loos, E. A. Croft, and J. Garland. Design and evaluation of a touch-centered calming interaction

with a social robot. *Trans Affective Computing*, PP(99), 2015. \rightarrow pages ii, 6, 9, 17, 39, 46, 81, 83, 93

- [89] H. Sharp. Interaction design. John Wiley & Sons, 2003. \rightarrow pages ii, 9, 76, 94
- [90] T. Shibata. Ubiquitous surface tactile sensor. In *Robotics and Automation*, 2004. TExCRA'04. First IEEE Technical Exhibition Based Conference on, pages 5–6. IEEE, 2004. → pages 5, 6, 7
- [91] T. Shibata and K. Wada. Robot therapy: a new approach for mental healthcare of the elderly-a mini-review. *Gerontology*, 57(4):378–386, 2010. → pages 18
- [92] K. B. Shimoga. Finger force and touch feedback issues in dexterous telemanipulation. In *Proc Intelligent Robotic Systems for Space Exploration*, pages 159–178, 1992. → pages 10, 17, 58, 92
- [93] D. Silvera-Tawil, D. Rye, and M. Velonaki. Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *Intl J of Social Robotics*, 6(4):489–505, 2014. → pages 1, 18, 48
- [94] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf. Design of a therapeutic robotic companion for relational, affective touch. In *ROMAN Workshop on Robot & Human Interactive Communication*, pages 408–415, 2005. → pages 5, 18
- [95] W. D. Stiehl, C. Breazeal, K.-H. Han, J. Lieberman, L. Lalla, A. Maymin, J. Salinas, D. Fuentes, R. Toscano, C. H. Tong, et al. The huggable: a therapeutic robotic companion for relational, affective touch. In ACM SIGGRAPH 2006 emerging technologies, page 15. ACM, 2006. → pages 5, 38
- [96] W. D. Stiehl, J. K. Lee, C. Breazeal, M. Nalin, A. Morandi, and A. Sanna. The huggable: a platform for research in robotic companions for pediatric care. In *Proceedings of the 8th International Conference on interaction Design and Children*, pages 317–320. ACM, 2009. → pages 6, 10
- [97] T. Tamura, S. Yonemitsu, A. Itoh, D. Oikawa, A. Kawakami, Y. Higashi, T. Fujimooto, and K. Nakajima. Is an entertainment robot useful in the care of elderly people with severe dementia? *J Gerontology Series A: Biological Sciences and Medical Sciences*, 59(1):M83–M85, 2004. → pages 18

- [98] D. S. Tawil, D. Rye, and M. Velonaki. Touch modality interpretation for an eit-based sensitive skin. In *IEEE Conf on Robotics and Automation (ICRA)*, pages 3770–3776, 2011. → pages 27
- [99] F. Thomas and O. Johnston. *Disney animation: The illusion of life*. Abbeville Press, 1981. → pages 82
- [100] J. Ulmen and M. Cutkosky. A robust, low-cost and low-noise artificial skin for human-friendly robots. In *Proc. IEEE Robotics and Automation* (*ICRA*), pages 4836–4841, 2010. → pages 19
- [101] J. D. Velasquez. An emotion-based approach to robotics. In Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on, volume 1, pages 235–240. IEEE, 1999. → pages 83
- [102] K. Wada and T. Shibata. Living with seal robotsits sociopsychological and physiological influences on the elderly at a care house. *Robotics, IEEE Transactions on*, 23(5):972–980, 2007. → pages 38
- [103] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In *Robotics and Automation*, 2005, pages 2785–2790. IEEE, 2005. → pages 1, 83
- [104] K. Wada, Y. Ikeda, K. Inoue, and R. Uehara. Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot paro. In *ROMAN*, pages 533–538, 2010. → pages ii, 6, 18
- [105] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988. → pages 43
- [106] P. Wells. The animated bestiary: animals, cartoons, and culture. Rutgers University Press, 2008. → pages 82
- [107] S. Yohanan and K. MacLean. The haptic creature project: Social human-robot interaction through affective touch. In *Proc AISB 2008*, volume 1, pages 7–11. Citeseer, 2008. → pages 83
- [108] S. Yohanan and K. E. MacLean. A tool to study affective touch. In CHI'09 Extended Abstracts on Human Factors in Computing Systems, pages 4153–4158. ACM, 2009. → pages 5, 6, 81

- [109] S. Yohanan and K. E. MacLean. Design and assessment of the haptic creature's affect display. In *Proceedings of the 6th international conference* on Human-robot interaction, pages 473–480. ACM, 2011. → pages 13, 83
- [110] S. Yohanan and K. E. MacLean. The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *International Journal of Social Robotics*, 4(2):163–180, 2012. → pages ii, x, 1, 2, 15, 19, 21, 38, 44, 53, 89, 94

Appendix A

Supporting Materials

A.1 Study Forms

A.1.1 Consent Form



Department of Computer Science 2366 Main Mall Vancouver, B.C. Canada V6T 1Z4 tel: (604) 822-3061 fax: (604) 822-4231 Project Title: Investigation of Interactive Affective Touch

Principal Investigator: Karon MacLean, Professor, Dept. of Computer Science, 604-822-8169 Co-Investigator: Xi Laura Cang, MSc Student, Dept. of Computer Science, 604-827-3982

The purpose of this study is to gather feedback to inform the interaction design of an haptic zoomorphic robot. We may ask you to interact with one or more touch sensitive surfaces mounted on a variety of stationary and/or moving objects. . We may also ask you to interact with an interface for controlling small robots, and ask you to create, manipulate, or describe the motions and perceived emotional content.. We may ask you to talk about your experiences with animals and pets. This study is part of a graduate student research project.

You may refuse or skip any task or question without affecting your reimbursement.

REIMBURSEMENT:	We are very grateful for your participation. You will receive monetary compensation of \$10 for this session.				
TIME COMMITMENT:	1×1 hour session				
RISKS & BENEFITS:	This experiment contains no more risk than everyday computer use or commercially available actuated toys. There are no direct benefits to participants beyond compensation.				
CONFIDENTIALITY:	You will not be identified by name in data gathered from this experiment Science account accessible only to the excerpts will be edited to remove ide not limited to obscuring face and/or publication unless permission is exp	any study i will be store he experime entifying inf voice) and licitly given	reports. Any id ed in a secure enters. Video d formation (incl will not be use below.	dentifiable Computer or audio luding but ed in	
AUDIO/VIDEO RELEASE:	SE: You may be asked for audio or video to be recorded during this sessio You are free to say no without affecting your reimbursement.				
	I agree to have AUDIO recorded:		Yes	□ No	
	I agree to have VIDEO recorded: I agree to have ANONYMIZED VII	DEO OR A	Yes UDIO EXCEF	□ No RPTS	
	presented in publications:	□ Yes	🗆 No		

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study. Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study. Any questions about the study can be directed to Laura Cang, cang@cs.ubc.ca.

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the

Version 2.0 / June 6, 2016 / Page 1 of 2

UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

You hereby CONSENT to participate and acknowledge RECEIPT of a copy of the consent form:

PRINTED NAME _____ DATE ____

SIGNATURE _____

Version 2.0 / June 6, 2016 / Page 2 of 2

A.1.2 Call for Participation Form

THE UNIVERSITY OF BRITISH COLUMBIA



Department of Computer Science 201-2366 Main Mall Vancouver, B.C. Canada V6T 1Z4 tel: (604) 822-3061 fax: (604) 822-4231

Investigation of Interactive Affective Touch

Principal Investigator: Karon MacLean, Professor, Dept. of Computer Science, 604-822-8169 Co-Investigator: Laura Cang, MSc Student, Dept. of Computer Science, 604-827-3982 Version 2.0 / June 6, 2016

The following message will be used to recruit participants for our study. We will distribute this message using some or all of the following methods:

- Emailing the recruitment message to mailing lists maintained by the Computer Science department or our research group, such as a list of department graduate students (often used for this kind of purpose) and a list of persons who have expressed an interest in being study participants.
- Uploading the recruitment message as an online posting, on craigslist.ca or facebook.
- Physical postings in public areas.
- Email and word-of-mouth when conducting purposeful sampling.

From: Laura Cang

Subject: Call for Study Participants - \$10 for Interactive Affective Touch

The Sensory, Perception, and Interaction (SPIN) Research Group in the UBC Dept. of Computer Science is looking for participants for a study investigating the sensing, design, and interpretation of emotive interactions with a small furry robot and/or other household objects. You will be compensated \$10 for your participation in a single 1-hour session.

We may ask you to talk about your experiences with household pets and other animals as well as other emotion-rich stories or memories. We may ask you to interact with touch sensitive surfaces or one or more robots that may produce any number of sounds, motions, and/or vibrations. We may also ask you to interact with a device for controlling these robots, and ask you to create, manipulate, or describe haptic (touch) sensations. Your touch and eye-gaze interactions may be recorded.

Please visit <URL> or contact me to sign-up for the study. You may also contact me if you have any questions.

Laura Cang MSc Student, UBC Computer Science cang@cs.ubc.ca

A.2 Participant Response Forms

A.2.1 Gesture Study Demographic Questionnaire

Participant ID

- 1. What is your age?
- 2. What is your gender?
- 3. What languages do you speak?
 - a. primary b. secondary c. other
- 4. How often do you interact with a pet?

1 0 2 0 3 0 4 0 5 Rarely Frequently

Participant ID

Do you have a pet?

If yes, what kind of pet is it? Describe your relationship with your pet.

[] loving [] indifferent [] neutral [] frustrating [] hateful

5. What kind of emotions are you trying to convey when you _____ on a pet, if any?

- a. Constant (laying your hand without motion)
- b. No touch
- c. Pat
- d. Rub
- e. Scratch
- f. Stroke
- g. Tickle
- 6. What did you think of the form that you were touching?

Does it resemble an animal, if so, what kind? If not, what features would it need to have to do so?

7. Anything to add?

A.2.2 Affective Rating Form

1. How real or genuine was the emotion you experienced when telling the story?

0	1	2	3	4	5	6	7	8	9	10
(totally a	rtificial)			(mo	derately i	real)		(com	pletely ge	enuine)

2. How did you feel when telling the story? Please check the corresponding box below.



A.2.3 Robot Behaviour Interview Script

PARTICIPANT ID: _____ DATE: _____

Semi-Structured Interview Script

Experimenter: Thank you for participating in our study. We would like to ask a few questions about your impressions of the robot display. If you require clarification or are uncomfortable for any reason, feel free to interrupt at any time.

Form-factor Impressions:

1. What are your initial impressions of each robot? How would you describe these forms (e.g., machine, animal, cartoon, ...)?

- 2. Where might they have come from?
- 3. What does it mean when they are not moving?
- 4. How you might use these robots?
- 5. Which robot seems to like you more and why?

Comparative emotional clarity of the robot displays:

6. Do you think there were differences in how the robots were able to express their feelings?

7. Did one seem more dramatic than the other?

Which robot would you	pick for expressing each emo	otion?	
Stressed	Excited	Depressed	Relaxed

Pet Preferences:

9. Have you had any experience with household pets or other domesticated animals?

- If yes, tell me about your pet or animal.

- If not, do you want a pet, why or why not?

10. Do you have any questions or comments for us?

Version 0.1 / 13 October, 2015 / Page 1 of 1