

# Non-Invasive Wearable Sensing Systems for Continuous Health Monitoring and Long-Term Behavior Modeling

by

**Michael Sung**

B.S., Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1997  
M.Eng., Electrical Engineering and Computer Science, Massachusetts Institute of Technology 1998

Submitted to the Department of Electrical Engineering and Computer Science,  
School of Engineering,

in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in Electrical Engineering and Computer Science  
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October 2005 *February 2006*

© 2005 Massachusetts Institute of Technology. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others the right to do so.

Author .....

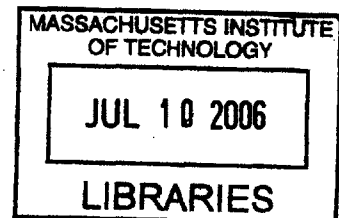
Michael Sung  
Department of Electrical Engineering and Computer Science  
October 12, 2005

Certified by .....

Alex P. Pentland  
Professor of Media Arts and Sciences  
Thesis Supervisor

Accepted by .....

Arthur C. Smith  
Chairman, Departmental Committee on Graduate Students  
Program in Electrical Engineering and Computer Science



ARCHIVES

# **Non-Invasive Wearable Sensing Systems for Continuous Health Monitoring and Long-Term Behavior Modeling**

by

**Michael Sung**

Submitted to the Department of Electrical Engineering and Computer Science,  
School of Engineering,  
on October 12, 2005, in partial fulfillment of the  
requirements for the degree of  
DOCTOR OF PHILOSOPHY

## **Abstract**

Deploying new healthcare technologies for proactive health and elder care will become a major priority over the next decade, as medical care systems worldwide become strained by the aging populations. This thesis presents LiveNet, a distributed mobile system based on low-cost commodity hardware that can be deployed for a variety of healthcare applications. LiveNet embodies a flexible infrastructure platform intended for long-term ambulatory health monitoring with real-time data streaming and context classification capabilities. Using LiveNet, we are able to continuously monitor a wide range of physiological signals together with the user's activity and context, to develop a personalized, data-rich health profile of a user over time.

Most clinical sensing technologies that exist have focused on accuracy and reliability, at the expense of cost-effectiveness, burden on the patient, and portability. Future proactive health technologies, on the other hand, must be affordable, unobtrusive, and non-invasive if the general population is going to adopt them. In this thesis, we focus on the potential of using features derived from minimally invasive physiological and contextual sensors such as motion, speech, heart rate, skin conductance, and temperature/heat flux that can be used in combination with mobile technology to create powerful context-aware systems that are transparent to the user. In many cases, these non-invasive sensing technologies can completely replace more invasive diagnostic sensing for applications in long-term monitoring, behavior and physiology trending, and real-time proactive feedback and alert systems.

Non-invasive sensing technologies are particularly important in ambulatory and continuous monitoring applications, where more cumbersome sensing equipment that is typically found in medical and clinical research settings is not usable. The research in this thesis demonstrates that it is possible to use simple non-invasive physiological and contextual sensing using the LiveNet system to accurately classify a variety of physiological conditions. We demonstrate that non-invasive sensing can be correlated to a variety of important physiological and behavioral phenomenon, and thus can serve as substitutes to more invasive and unwieldy forms of medical monitoring devices while still providing a high level of diagnostic power.

From this foundation, the LiveNet system is deployed in a number of studies to quantify physiological and contextual state. First, a number of classifiers for important health and general contextual cues such as activity state and stress level are developed from basic non-invasive physiological sensing. We then demonstrate that the LiveNet system can be used to develop systems that can classify clinically significant physiological and pathological conditions and that are robust in the presence of noise, motion artifacts, and other adverse conditions found in real-world situations. This is highlighted in a cold exposure and core body temperature study in collaboration with the U.S. Army Research Institute of Environmental Medicine. In this study, we show that it is possible to develop real-time implementations of these classifiers for proactive health monitors that can provide instantaneous feedback relevant in soldier monitoring applications.

This thesis also demonstrates that the LiveNet platform can be used for long-term continuous monitoring applications to study physiological trends that vary slowly with time. In a clinical study with the Psychiatry Department at the Massachusetts General Hospital, the LiveNet platform is used to continuously monitor clinically depressed patients during their stays on an in-patient ward for treatment. We show that we can accurately correlate physiology and behavior to depression state, as well as to track changes in depression state over time through the course of treatment. This study demonstrates how long-term physiology and behavioral changes can be captured to objectively measure medical treatment and medication efficacy. In another long-term monitoring study, the LiveNet platform is used to collect data on people's everyday behavior as they go through daily life. By collecting long-term behavioral data, we demonstrate the possibility of modeling and predicting high-level behavior using simple physiologic and contextual information derived solely from ambulatory mobile sensing technology.

Thesis Supervisor: Alex P. Pentland

Title: Toshiba Professor of Media Arts and Sciences, MIT

# **Non-Invasive Wearable Sensing Systems for Continuous Health Monitoring and Long-Term Behavior Modeling**

by

**Michael Sung**

Thesis Committee:

Advisor: .....

Prof. Alex (Sandy) Pentland  
Department of Media Arts and Sciences  
Massachusetts Institute of Technology

Thesis Reader .....

Prof. John Guttag  
Department of Electrical Engineering and Computer Science  
Massachusetts Institute of Technology

Thesis Reader .....

Prof. Peter Szolovits  
Department of Electrical Engineering and Computer Science  
Massachusetts Institute of Technology

Thesis Reader .....

Dr. Carl Marci  
Director of Social Neuroscience  
Massachusetts General Hospital/Harvard Medical School



# Acknowledgments

First and foremost, I would like to thank my parents and sister, who have unconditionally and tirelessly supported me through my education through the first quarter of my life. My dad is truly my role model, my mom is the greatest cheerleader a son could ever have, and my sister is, well, my only little sister.

To Sandy Pentland, thanks for being absolutely the most ideal advisor I could possibly have asked for. You are a true visionary with a stature in both industry and academia I can only hope to emulate someday. I am deeply thankful for all the guidance you have shown me while I have defined my research niche. I am very much looking forward to collaborating with you for a long time to come.

To my committee members, Peter Szolovits and John Guttag, thanks for sharing an infectious enthusiasm for my research and providing the feedback I needed to direct my research. I would also like to thank Carl Marci, who has become a colleague, mentor, and most importantly, friend. I am truly honored to be able to have such a prestigious and well-rounded doctoral committee for my thesis.

To Jon Gips, my counterpart running the Wearable Computing Group...the yin to my yang, we have become a formidable pair working together over the years (especially with a sponsor meeting or conference deadline looming). Nothing that has come out of the group would have been possible without your countless selfless debugging sessions. I hope you're ready to launch Project High-Note and Project Dicktator with me... genius, baby, genius.

To Nathan Eagle, it's probably a good thing we never got around to turning our office into the disco playland of our dreams...we both made it! Looking forward to meeting up with you in the upcoming year to begin my Life Masters training, which obviously is a lifelong endeavor (much as pursuing this Ph.D. has seemed one).

To Joost Bonsen, you have an amazing talent to bring people together and even more uncanny ability to rattle off an endless stream of new buzzwords of which I was no less a beneficiary of than just about everybody else within the MIT community. No doubt you will do great things in life, and I hope that I can remain a good friend and colleague and see you through it all.

To Jane Liaw, who has been so cool over the years and for being the proverbial peegoo I could lean on. If I spent just a hundredth of the time I spent talking to you online actually working on research, I would have been able to write 5 theses, only. Thanks for being my most favorite person to waste time and share life with.

To Barney Monteith, seems like I've done just about everything with you over the last few years. You really opened my eyes to the world of possibility with your make-things-happen attitude. Thanks for the patience in keeping ADS afloat while I finished my thesis. I can't wait to take over the world with you, man.

To all my other friends who have all either been supportive of me or tried to distract me. Mar, you're basically the only one on the supportive side of the fence...:-) You continued to believe in me even when I doubted myself. Your support has meant a great deal to me. To my crew: Taz, my best of buds, looking forward to all of the Epic venues yet to come.... Dow, Giddy-up! Your turn to become a doctor. Chol, I'm glad you moved to NJ cuz I'd still be at your place trying to do that 3-ramen century combo if you didn't leave. Pauly, I still want the front seat next time. Jay, thanks for schooling me on some of the finer things in life, especially how to hit the high-notes singing Patience at our countless Notoban sessions...I'll finally be back in SF so we can resume our hijinx. To Jamie, your enthusiasm for the weird geeky thing attached to my arm is shocking...you rock, Fashion. Big props to every single one of you who made my journey through grad school tolerable.

To the rest of the Pentlandians, Silvia Jimenez, Juan Carlos, Ron Caneel, Wen Dong, Anmol Madan, all who really made the lab an enjoyable place to hang out. I'm looking forward to getting kicked out of the CBC when the time comes for each of you.

I would also like to acknowledge the rest of the Borglab crew past and present (Vadim Gerasimov, Rich DeVaul, Chris Elledge, Dan Barkalow, Edward Keyes, Bradley Rhodes, Thad Starner, Dkap, others). Obviously, many of the things that have been done in this thesis have been possible only because of the path blazed by the people before me. I am very honored to be able to list myself as a peer and now alumnus of the Wearable Computing Group along with the rest of you.

To all my undergraduate research advisees, Ben Maron, Rimi Ganguly, John Pavlish, Xingpeng Huang, Xavid Pretzer, Justin Cannon, Pramook Khungurn, Albert Sun, Cemal Akcaba, Jeffrey Huang, Kathen Chen, Hyunsuk Kim, Shirli Li, Jennifer Yeh, Jeff Pan, Shaun Foley, Leo, and Hiba Awad, I hope you got as much out of working with me as I with you, and that all of you learned something useful along the way.

I also want to thank the rest of the people I've been lucky enough to work with over the years (Carl Marci, John Castellani, John Ames, Torsten Linz, Stephen Intille, Spencer Lynn, Dr. Berg, and all the other people I am forgetting now) on the interesting discussions and research collaborations that eventually paved the way to the studies central to my thesis.

As a transplant originally from LCS, the Media Lab community has really been something out of this world. I am very glad I had the chance to work here, and thank all the ML faculty and colleagues that have collectively opened my eyes to the exciting possibilities of technology and design.

Well, I have been here at MIT long enough and it's time to move on, but I do so knowing I will look back upon my memories here from what has been my home for the last 12 years with warmth and fondness.

# Table of Contents

<b>ABSTRACT</b> .....	<b>I</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>IV</b>
<b>TABLE OF CONTENTS</b> .....	<b>VII</b>
<b>LIST OF FIGURES</b> .....	<b>XIV</b>
<b>LIST OF TABLES</b> .....	<b>XXIII</b>
<b>LIST OF TABLES</b> .....	<b>XXIII</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
1.1 THESIS CONTRIBUTIONS .....	9
1.2 THESIS ROADMAP.....	10
<b>CHAPTER 2 BACKGROUND</b> .....	<b>14</b>
2.1 HISTORY OF WEARABLE HEALTH TECHNOLOGY.....	14
2.2 HISTORY OF TELEMEDICINE .....	17
2.3 GOLD-STANDARD VS. NON-INVASIVE SENSING.....	18
2.4 POST PROCESSING VS. REAL-TIME FEEDBACK.....	20
2.5 EVENT REPORTING VS. CONTINUOUS MONITORING .....	22
2.6 SHORT-TERM VS. LONG-TERM MONITORING .....	24
2.7 REACTIVE VS. PROACTIVE TECHNOLOGIES .....	25
2.8 PERVASIVE INFRASTRUCTURE VS. MOBILE TECHNOLOGY.....	27
2.9 CENTRALIZED VS. DISTRIBUTED COMPUTING .....	28
2.10 INDIVIDUALIZED VS. GROUP APPLICATIONS.....	30
<b>CHAPTER 3 WEARABLE AND SENSING INFRASTRUCTURE</b> .....	<b>32</b>

3.1 SYSTEM HARDWARE TECHNOLOGY .....	36
3.2 SAK2 DATA ACQUISITION AND SENSOR HUB BOARD.....	38
<i>SAK2 Specifications</i> .....	39
<i>SAK2 Firmware</i> .....	41
3.3 PHYSIOLOGIC AND CONTEXTUAL SENSING TECHNOLOGY .....	41
<i>BioSense Specifications</i> .....	42
<i>Other Sensors/Peripherals</i> .....	44
3.4 HUMAN FACTORS RESEARCH .....	45
3.5 SOFTWARE SYSTEM INFRASTRUCTURE.....	48
<i>The Enchantment Whiteboard</i> .....	50
<i>Enchantment Whiteboard server</i> .....	51
<i>Enchantment Signal System</i> .....	51
<i>MIThril Real-Time Context Engine</i> .....	53
3.6 LIVE.NET APPLICATION DEVELOPMENT.....	54
<i>BioRecord: GUI-Based Multimodal Sensor Acquisition and Annotation</i> .....	55
<i>Quest: Configurable Experience Sampling Application</i> .....	56
<i>PatientMonitor: Robust Automated Monitoring System</i> .....	57
<i>PatientTracker: Location Beacon Person Tracker</i> .....	58
<i>Real-Time Physiometric Displays</i> .....	60
<i>VitaMon: Low-Cost Critical Health Monitors</i> .....	61
<i>BorgViewer: Distributed Real-Time Data Communication System</i> .....	62

3.7 LIVE <span>NET</span> COLLABORATIONS.....	64
<i>Physiology and Violence in Games Study, Harvard Medical School</i> .....	64
<i>ECG Shirt/Non-invasive Sensing Technologies, Franhauser Institute</i> .....	64
<i>Pervasive Healthcare Applications, National Taiwan University</i> .....	65
<i>Eldercare Monitoring, British Telecom</i> .....	65
<i>Parkinson's Disease Study, Universidad Polit3cnica de Madrid</i> .....	66
<i>Epilepsy Study, University of Rochester Center for Future Health</i> .....	66
<i>Soldier Health Monitoring, U.S Army ARIEM Natick Laboratories</i> .....	67
<i>Depression Study, Massachusetts General Hospital</i> .....	68
<b>CHAPTER 4 NON-INVASIVE FEATURES AND MODELING</b> .....	<b>69</b>
4.1 GENERIC PHYSIOLOGY FEATURES .....	69
4.2 ACCELEROMETRY/MOTION .....	74
4.3 SOCIAL INTERACTION AND SPEECH FEATURES.....	75
4.4 TEMPERATURE/HEAT FLUX.....	81
4.5 HEART RATE AND HEART RATE VARIABILITY .....	82
4.6 SKIN CONDUCTANCE RESPONSE.....	87
4.7 BEHAVIORAL FEATURES.....	90
<i>Sleeping Behavior</i> .....	90
<i>Activity Behavior</i> .....	91
<i>Social Behavior</i> .....	91
4.8 CORRELATING NON-INVASIVE SENSING TO CLINICAL MEASURES .....	92

<i>Heart Rate vs. ECG</i> .....	92
<i>Heart Rate vs. Respiration</i> .....	93
<i>Skin conductance vs. Gold Standard</i> .....	93
4.9 DEALING WITH MULTIMODAL SENSING AND FEATURE REDUCTION.....	94
<i>Correlating Features</i> .....	95
<i>Principal Component Analysis</i> .....	96
4.10 MODEL SELECTION AND CLASSIFICATION TECHNIQUES.....	99
<i>Multiple Linear Regression</i> .....	99
<i>Stepwise Regression</i> .....	100
<i>Gaussian Mixture Models</i> .....	100
<i>Hidden Markov Models</i> .....	102
<b>CHAPTER 5 NON-INVASIVE CONTEXT CLASSIFICATION.....</b>	<b>107</b>
5.1 GENERAL ACTIVITY CONTEXT CLASSIFICATION .....	107
<i>General Activity Context Experimental Methodology</i> .....	108
<i>General Activity Context Analysis</i> .....	109
<i>General Activity Context Results</i> .....	110
<i>General Activity Context Discussion</i> .....	111
5.2 POKERMETRICS STRESS AND PHYSIOLOGY STUDY .....	113
<i>Background on the Physiology of Human Intent and Affect</i> .....	114
<i>PokerMetrics Human Subjects Approval</i> .....	115
<i>PokerMetrics Subject Selection</i> .....	115

<i>PokerMetrics Experimental Methodology</i> .....	115
<i>PokerMetrics Subjective and Objective Outcome Data</i> .....	116
<i>PokerMetrics Data Analysis</i> .....	117
<i>Creating Stress and Lie Detectors from Physiology</i> .....	123
<i>PokerMetrics Discussion</i> .....	124
<b>CHAPTER 6 REAL-TIME CLINICAL HEALTH CLASSIFICATION</b> .....	<b>126</b>
6.1 ARIEM COLD EXPOSURE/HYPOTHERMIA STUDY .....	126
6.2 ARIEM EXPERIMENTAL METHODOLOGY .....	127
<i>ARIEM Study Human Subjects Approval</i> .....	127
<i>ARIEM Study Subject Selection</i> .....	128
<i>ARIEM Protocol</i> .....	128
6.3 ARIEM SENSING AND DATA COLLECTION.....	129
6.4 ARIEM DATA ANALYSIS AND FEATURE SELECTION .....	130
6.5 ARIEM MODEL SELECTION, TRAINING, AND CLASSIFICATION.....	133
6.6 ARIEM REAL-TIME WEARABLE IMPLEMENTATION.....	135
6.7 ARIEM DISCUSSION AND FUTURE DIRECTIONS.....	136
<b>CHAPTER 7 CONTINUOUS MONITORING &amp; CLINICAL TRENDING</b> .....	<b>138</b>
7.1 MGH DEPRESSION STATE AND ECT TREATMENT STUDY.....	138
7.2 BACKGROUND DEPRESSION PHYSIOLOGY RESEARCH .....	140
7.3 MGH EXPERIMENTAL PROTOCOL .....	144
<i>MGH Human Subjects Approval</i> .....	144



<i>MGH Subject Selection</i> .....	144
<i>MGH Subject Recruitment</i> .....	145
<i>MGH Experimental Procedures</i> .....	146
7.4 MGH SENSING AND DATA COLLECTION.....	147
<i>Physiology and Behavioral Measures</i> .....	148
<i>Clinical Psychiatric Outcome Measures</i> .....	149
<i>Self-Report Emotional Rating Surveys</i> .....	150
7.5 MGH DATA ANALYSIS.....	151
<i>MGH Data Preprocessing</i> .....	152
<i>Raw Emotional Rating Survey Data</i> .....	154
<i>Decomposing the Emotion Rating Data</i> .....	157
<i>Creating a Composite Depression Emotion Factor</i> .....	161
<i>Physiology and Behavioral Data</i> .....	163
<i>Depression Regression Analysis</i> .....	165
<i>Feature Dimensionality Reduction Analysis</i> .....	168
<i>Clinical Outcome Trending</i> .....	171
<i>ECT vs. non-ECT days</i> .....	173
<i>Correlating Behavioral Features</i> .....	175
<i>Longitudinal Subject Results</i> .....	176
<i>Toward Creating a Universal Depression Meter</i> .....	182
7.6 MGH STUDY DISCUSSION.....	183

<b>CHAPTER 8 LONG-TERM BEHAVIORAL MODELING.....</b>	<b>187</b>
8.1 LIFESCAPE STUDY BACKGROUND .....	187
8.2 LIFESCAPE EXPERIMENTAL PROTOCOL .....	190
<i>Behavior, Location, and Interaction Annotation Log</i> .....	190
<i>Cell Tower Location IDs and Bluetooth Interaction Logs</i> .....	191
<i>The Lifescape Dataset</i> .....	191
8.3 LIFESCAPE BEHAVIOR MODEL .....	192
8.4 LIFESCAPE RESULTS .....	194
8.5 LIFESCAPE DISCUSSION .....	196
<b>CHAPTER 9 FUTURE DIRECTIONS.....</b>	<b>199</b>
9.1 REAL-TIME FEEDBACK APPLICATIONS.....	199
9.2 DISTRIBUTED PEER-TO-PEER HEALTHCARE APPLICATIONS .....	200
<b>CHAPTER 10 CONCLUSIONS .....</b>	<b>203</b>
10.1 PROTOTYPING TECHNOLOGIES FOR FUTURE HEALTHCARE APPLICATIONS .....	203
10.2 DATA-MINING INDIVIDUALS: EXPLORING HUMAN PHYSIOLOGY AND BEHAVIOR .....	205
10.3 PRIVACY IMPLICATIONS AND THE PRICE OF PRIVACY .....	206
10.4 LIMITATIONS TO LIVE <span style="font-variant: small-caps;">NET</span> .....	208
10.5 GLIMPING INTO THE FUTURE OF HEALTHCARE.....	209
<b>REFERENCES .....</b>	<b>212</b>

# List of Figures

FIGURE 1: LIVENET WEARABLE PERFORMING REAL-TIME FFT ANALYSIS AND ACTIVITY CLASSIFICATION ON ACCELEROMETER DATA, VISUALIZING THE RESULTS, AND WIRELESSLY STREAMING REAL-TIME ECG/SKIN CONDUCTANCE/TEMPERATURE AND CLASSIFICATION RESULTS TO A REMOTE COMPUTER WITH A PROJECTION DISPLAY AS WELL AS PEER LIVENET SYSTEMS (WHICH MAY BE LOCATED ANYWHERE IN THE WORLD). .....3

FIGURE 2: HIERARCHICAL LIVENET ARCHITECTURE. EACH LIVENET SYSTEM IS COMPRISED OF A MOBILE DEVICE (WITH LOCAL PROCESSING, STORAGE, AND WIRELESS CAPABILITIES), THE SAK2 SENSOR HUB, AND A CONFIGURABLE SENSOR NETWORK. THE LIVENET ARCHITECTURE CAN BE CONFIGURED FOR BOTH TRADITIONAL CLIENT-SERVER APPLICATIONS AS WELL AS PEER-TO-PEER COMMUNICATIONS FOR DISTRIBUTED APPLICATIONS INVOLVING MULTIPLE SYSTEMS. DATA STREAMS CAN BE SUBSCRIBED TO BY BACKEND SERVERS (FOR MORE INTENSIVE PROCESSING THAT CANNOT BE HANDLED LOCALLY), DISPLAYS, AND OTHER ENDPOINTS. ....35

FIGURE 3: LIVENET SYSTEM, COMPOSED OF THE ZAURUS PDA (TOP LEFT), WITH SAK2 DATA ACQUISITION/SENSOR HUB AND BIOSENSE PHYSIOLOGICAL SENSING BOARD (MIDDLE), BATTERY SOURCE (TOP RIGHT), SENSOR BUS HUB (LOWER RIGHT), 3D ACCELEROMETER BOARD (MIDDLE LEFT), AND WMSAD MULTISENSOR BOARD (LOWER LEFT). .....37

FIGURE 4: SAK2 DATA ACQUISITION AND SENSOR HUB BOARD (FRONT/BACK). THE SAK2 IS THE PRIMARY INTERFACE FOR THE LIVENET PLATFORM BETWEEN THE EMBEDDED MOBILE DEVICE AND THE CUSTOMIZABLE PLUG-AND-PLAY SENSOR NETWORK AND A SUPPORTS A VARIETY OF WIRELESS AND WIRED COMMUNICATION PROTOCOLS. THE SAK2 HAS AN ON-BOARD POWER REGULATOR CAPABLE OF PROVIDING POWER TO THE LIVENET SYSTEM COMPONENTS AS WELL AS THE SENSOR NETWORK DURING SUSTAINED MONITORING APPLICATIONS. THE SAK2 CAN ALSO BE USED FOR STAND-ALONE DATA ACQUISITION APPLICATIONS. ....39

FIGURE 5: BIOSENSE DAUGHTERBOARD (FRONT/BACK). THE BIOSENSE BOARD IS DESIGNED TO WORK DIRECTLY WITH THE SAK2 SENSOR HUB, PROVIDING 3D MOTION, SKIN CONDUCTANCE, TEMPERATURE, ECG/EMG SENSING AS WELL PROTOTYPING AMPLIFICATION CIRCUITRY FOR ANALOG SIGNALS. THE BIOSENSE CAN ALSO ACT AS A BRIDGE CONVERTER TO ATTACH OTHER SERIAL-BASED 3<sup>RD</sup> PARTY SENSORS TO THE SENSOR NETWORK. ....42

FIGURE 6: OPEN AND SIDE PROFILES OF A LIVENET HOLSTER PACKAGE (TOP) AND CRADLE ASSEMBLY FOR EASY DOCKING OF THE ZAURUS TO THE SAK2/BIOSENSE SUBSYSTEM (BOTTOM). THIS SYSTEM ALLOWS THE ZAURUS TO BE REMOVED FROM THE HOLSTER FOR INTERACTION APPLICATIONS WHILE SIMULTANEOUSLY ALLOWING THE SAK2 SENSOR HUB TO BUFFER AND STORE SENSOR DATA, WHICH CAN BE SYNCED ONCE THE DEVICE IS RE-DOCKED WITH THE CRADLE.....47

FIGURE 7: WEARABLE OPTIONS INCLUDING A BANDOLIER-STYLE SYSTEM (LEFT) AND A ‘FANNY-PACK’ WAIST STRAP SYSTEM (RIGHT). THE BANDOLIER SYSTEM WAS DESIGNED AS A COMPACT FORM FACTOR THAT ALLOWS THE PDA DEVICE TO BE QUICKLY ACCESSED FOR INTERACTION APPLICATIONS. SENSORS SUCH AS THE AUDIO MICROPHONE AND IR BEACONS CAN BE CONVENIENTLY ATTACHED TO THE STRAP ON THE CHEST. THE WAIST STRAP SYSTEM WAS MORE COMFORTABLE FOR WOMEN AND MORE CONVENIENT FOR LONG-TERM WEAR SUCH AS THE MGH DEPRESSION STUDY. ....48

FIGURE 8: TYPICAL LIVENET APPLICATION SHOWING INTERACTION BETWEEN THE ENCHANTMENT SERVER, ENCHANTMENT WHITEBOARD, ENCHANTMENT SIGNAL SYSTEM, AND MITHRIL REAL-TIME CONTEXT ENGINE. PRODUCERS OF DATA (SUCH AS AN ACCELEROMETER OR MICROPHONE DEVICE) CAN PUBLISH UNL HANDLES TO A PARTICULAR TYPE OF DATA ON THE ENCHANTMENT WHITEBOARD, WHICH IS MODERATED BY THE ENCHANTMENT SERVER. THE ENCHANTMENT SIGNAL SYSTEM THE FORWARDS THESE DATA STREAMS, CALLED SIGNALS, (DEPICTED BY ARROWS) TO FEATURE EXTRACTION MODULES, WHICH THEN SUBSEQUENTLY POST THEIR OWN OUTPUT AS UNLS THAT OTHER MODULES DOWNSTREAM CAN SUBSCRIBE TO. A MODEL CAN THEN USE THE FEATURES EXTRACTED FOR CLASSIFICATION, AND THE CONTEXT RESULTS CAN THEN BE SENT TO THE FRONT-END AGENT AND UI MODULES.....52

FIGURE 9: PROCESS FLOW FOR DEVELOPING CONTEXT-AWARE APPLICATIONS. REAL-WORLD EVENTS CAN BE DETECTED USING THE SENSING INFRASTRUCTURE OF THE LIVENET SYSTEM. THIS DATA MUST FIRST BE PRE-PROCESSED TO EXTRACT USEFUL FEATURES AMENABLE FOR MODELING. THESE FEATURES CAN THEN BE USED AS INPUTS TO TRAIN THE MODELS THAT WE SELECT TO APPLY TO THE SITUATION OF INTEREST. THESE MODELS CAN THEN BE USED TO PROVIDE INFERENCES BY LIVENET APPLICATIONS TO DEVELOP CONTEXT-AWARE APPLICATIONS TO CLASSIFY THE CURRENT CONTEXT OF THE SYSTEM OR EVEN PREDICT THE FUTURE STATE OF THE SYSTEM.....54

FIGURE 10: BIORECORD APPLICATION SHOWING THE SENSOR SELECTION SCREEN (LEFT) AND THE TIMESTAMP AND ANNOTATION SCREEN (RIGHT). THIS GUI-BASED APPLICATION IS A FRONT END TO THE ENCHANTMENT INFRASTRUCTURE, ALLOWING RESEARCHERS TO USE TURN-KEY LIVENET SYSTEMS TO EASILY RECORD AND ANNOTATE A WIDE VARIETY OF SENSING DATA FOR EXPERIMENTS. ....55

FIGURE 11: QUEST APPLICATION SHOWING THE VISUAL ANALOG SCALE SURVEY QUESTION (LEFT) AND SUBJECT COMMENTS SECTIONS (RIGHT). THIS CUSTOMIZABLE SURVEY APPLICATION CAN BE PROGRAMMED TO AUTO-START BASED ON TIMED INTERVALS OR EVEN EVENT-SPECIFIC TRIGGERS FROM THE SENSOR NETWORK OR OTHER INPUTS. ....57

FIGURE 12: BLUETOOTH LOCATION BEACON AND ASSOCIATED HOUSING AND POWER SUPPLY. THESE BEACONS ALLOW THE LIVENET SYSTEM TO TRACK THE LOCATION OF A PATIENT WITHIN LIVING SPACES. THE LOCATION BEACONS ARE COMPLETELY ADJUSTABLE IN RANGE, WITH AN EFFECTIVE RADIUS BETWEEN TEN AND OVER SEVERAL HUNDRED FT. ....59

FIGURE 13: LIVENET SYSTEM PERFORMING REAL-TIME FFT ANALYSIS AND CLASSIFICATION ON ACCELEROMETER DATA, AND WIRELESSLY STREAMING REAL-TIME EKG/GSR/TEMPERATURE AND CLASSIFICATION RESULTS TO A REMOTE PROJECTOR. WE CAN EASILY IMAGINE A SYSTEM SUCH AS THIS BEING USED TO STREAM DISTRIBUTED DATA TO A VARIETY OF ENDPOINTS FOR REAL-TIME SOLDIER OR CRITICAL MEDICAL MONITORING SYSTEMS. ....61

FIGURE 14: VITAMON MONITOR AND SENSORS (LEFT) AND REAL-TIME CRITICAL PHYSIOLOGY FROM BEING STREAMED TO A REMOTE LAPTOP FOR STORAGE AND VISUALIZATION THE VITAMON IS A VERY LOW-COST (SUB \$30 IN VOLUMES) PHYSIOLOGICAL EVENT MONITOR CAPABLE OF REAL-TIME DATA STREAMING BASED ON THE SAK2 DATA ACQUISITION BOARD. ....62

FIGURE 15: A “BORG” WEARING A LIVENET SYSTEM USING THE BORGVIEWER APPLICATION TO STREAM RAW ACCELEROMETER AND ECG SIGNALS AS WELL AS TO DISPLAY THE REAL-TIME CONTEXT CLASSIFICATION INFORMATION. A USER RUNNING BORGVIEWER CAN INSTANTANEOUSLY VISUALIZE THE REAL-TIME DATA OF A BORG AND ALSO GET A SENSE OF ITS IMMEDIATE CONTEXT BASED ON THE REAL-TIME ACTIVITY CLASSIFIER AND PROXIMITY INFORMATION OF OTHER NEARBY BORGs BASED ON IR BEACONS. THIS DISTRIBUTED NETWORK OF LIVENET SYSTEMS IS PERFECT FOR COMMAND-AND-CONTROL AND OTHER GROUP-BASED APPLICATIONS. ....63

FIGURE 16: NON-INVASIVE HEART RATE CHEST STRAP FOR DETECTING INTER-BEAT INTERVAL OF THE HEART. THESE TYPES OF SENSORS ARE VERY ROBUST IN THE PRESENCE OF LARGE PHYSICAL MOTION AND HAVE BEEN SHOWN TO BE HIGHLY CORRELATED TO MORE INVASIVE ECG MONITORING EQUIPMENT. MOST OF THE HEART RATE FEATURES OF INTEREST (SUCH AS HEART RATE VARIABILITY MEASURES) DO NOT REQUIRE ECG MORPHOLOGY INFORMATION, SO A NON-INVASIVE CHEST STRAP IS JUST AS IDEAL FOR HEART RATE SENSING. ....83

FIGURE 17: STANDARD SKIN CONDUCTANCE ELECTRODES PLACEMENT ON THE BASE OF THE INDEX AND MIDDLE FINGERS. SKIN CONDUCTANCE IS A VERY USEFUL MEASURE FOR DETECTING STRESS, EMOTION, AND STARTLE RESPONSES IN INDIVIDUALS BASED ON CHANGES IN SWEAT GLAND ACTIVITY CONTROLLED BY THE SYMPATHETIC AND PARASYMPATHETIC BRANCHES OF THE AUTONOMIC NERVOUS SYSTEM.....88

FIGURE 18: RAW SKIN CONDUCTANCE SIGNAL TRACE FROM THE POKERMETRICS STUDY SHOWING A RESPONSE TO A NUMBER OF STRESSFUL SITUATIONS (TOP), NORMALIZED DIFFERENCE (MID TOP), FILTERED DIFFERENCE (MID BOTTOM) AND NEGATIVE (RED) AND POSITIVE (BLUE) THRESHOLDS (BOTTOM). THE THRESHOLDS CAN BE SET AT DIFFERENT TRIGGER LEVELS TO PROVIDE A GAUGE TO THE NUMBER OF VARIOUS-INTENSITY STRESS OR EMOTION RESPONSE PEAKS THAT OCCUR OVER A PERIOD OF TIME. ....90

FIGURE 19: LONG-TERM MULTIMODAL HEAT FLUX (RED), ACCELEROMETER (GREEN), AND SKIN CONDUCTANCE (BLACK) DATA OVER TWO TYPICAL DAYS OF AN INDIVIDUAL. BY SIMPLE INSPECTION OF THE RAW DATA SIGNALS, ONE CAN VISUALLY MAKE OUT SUCH CONTEXTS SUCH AS SLEEPING (PERIODS OF LOW VARIANCE IN THE ACCELEROMETER), HIGH ACTIVITY (DURING PERIODS WITH HIGH MOTION VARIANCE AND LARGE HEAT FLUX SPIKES), AND EVEN WHEN RIDING IN A CAR (POSITIONAL ACCELERATION WITH LOW HEAT FLUX). ....95

FIGURE 20: GRAPHICAL MODEL OF A HIDDEN MARKOV MODEL UNROLLED FOR 3 TIME SLICES. THE HIDDEN STATE (DENOTED BY  $Q$ ), DETERMINES THE OBSERVABLE OUTPUT ( $O$ ) WHICH WE ARE TRYING TO MODEL. WE CAN USE A GIVEN SEQUENCE OF OBSERVATIONS TO TRAIN A HMM, WHICH CAN THEN BE USED FOR INFERENCE TO PREDICT THE FUTURE BEHAVIOR OF THE SYSTEM. ....103

FIGURE 21: SUBJECT WEARING THE NON-INVASIVE SENSORS (HEART RATE CHEST STRAP, FINGER SKIN CONDUCTANCE ELECTRODES, AUDIO MICROPHONE, AND INTEGRATED MOTION/HEAT FLUX/SKIN CONDUCTANCE MONITOR ALONG WITH THE LIVE NET PLATFORM RUNNING BIORECORD DURING A SESSION OF THE POKERMETRICS STUDY. WE USE THE COLLECTED PHYSIOLOGY TO CORRELATE TO MANUALLY RECORDED SUBJECTIVE OUTCOMES SUCH AS STRESS LEVEL, INTEREST, AND BLUFFING. .116

FIGURE 22: SKIN CONDUCTANCE (TOP), HEART RATE (MIDDLE), AND AUDIO (BOTTOM) DATA FOR A STRESSFUL ALL-IN HAND (LEFT) VS. A NON-EVENTFUL HAND (RIGHT). AS CAN BE SEEN, THE ALL-IN PLAY MIDWAY THROUGH THE STRESSFUL HAND WAS ANTICIPATED BY THE BLUE PLAYER WHO INITIATES IT, SHOWN AS A SMALL RAMP IN THE SKIN CONDUCTANCE TRACE IN THE TOP LEFT GRAPH. THIS CAUSES A LARGE SPIKE IN THE SKIN CONDUCTANCE OF THE RED PLAYER A LITTLE BIT LATER. WE

CAN ALSO SEE THAT IN GENERAL THERE WAS INCREASED SKIN CONDUCTANCE, HEART RATE VARIABILITY, AND VOICE ACTIVITY IN THE STRESSFUL HAND RELATIVE TO THE NON-STRESSFUL HAND. .... 118

FIGURE 23: RAW SKIN CONDUCTANCE SIGNAL TRACE FROM THE POKERMETRICS STUDY SHOWING A RESPONSE TO A NUMBER OF STRESSFUL SITUATIONS (TOP), NORMALIZED DIFFERENCE (MID TOP), FILTERED DIFFERENCE (MID BOTTOM) AND NEGATIVE (RED) AND POSITIVE (BLUE) THRESHOLDS (BOTTOM). THE THRESHOLDS CAN BE SET AT DIFFERENT TRIGGER LEVELS TO PROVIDE A GAUGE TO THE NUMBER OF VARIOUS-INTENSITY STRESS OR EMOTION RESPONSE PEAKS THAT OCCUR OVER A PERIOD OF TIME. .... 120

FIGURE 24: LIVE NET SYSTEM USED FOR THE ARIEM STUDY WITH ACCELEROMETER AND HEART RATE SENSOR NETWORK. GIVEN THE STRENUOUS OPERATING ENVIRONMENT OF THE PROTOCOL, THE SENSORS HAD TO BE COMPLETELY WATERPROOFED AND THE SYSTEM SHIELDED TO BE ROBUST IN THE PRESENCE OF LARGE AMOUNTS OF EM NOISE. THE LIVE NET SYSTEM IS CAPABLE OF NOT ONLY COLLECTING DATA, BUT IS ALSO USED TO PROTOTYPE THE REAL-TIME SHIVER DETECTION CLASSIFIER THAT WAS DEVELOPED. .... 129

FIGURE 25: CORE TEMPERATURE REGIMES OF A SUBJECT RESULTING FROM COLD EXPOSURE OVER A PERIOD OF AN HOUR AND A HALF. WE CAN SEE THAT THE CORE TEMPERATURE OF THE SUBJECT FALLS SLIGHTLY AS HE COOLS DOWN IN ROOM TEMPERATURE AIR IN THE BASELINE PERIOD (REGION 1). UPON ENTERING THE CHILLED WATER IN THE COLD PERIOD (REGION 2), THE SUBJECT BEGINS TO SHIVER SLIGHTLY AND THE CORE BODY TEMPERATURE IMMEDIATELY BEGINS TO DROP BEFORE LEVELING OFF AT THE END OF THE REGION. THE START OF THE COLDEST PERIOD (REGION 3) WAS PUNCTUATED WITH THE FIRST SIGNS OF INTENSE SHIVERING, AND THEREAFTER THE CORE BODY TEMPERATURE PLUMMETS UNTIL THE SUBJECT INDICATES HE CANNOT CONTINUE WITH THE PROTOCOL ANY LONGER. .... 130

FIGURE 26: ACCELEROMETER SPECTROGRAMS FOR CORE TEMPERATURE REGIMES IDENTIFIED IN FIGURE 25. WE CAN VISUALLY IDENTIFY THE DIFFERENT SPECTRAL CHARACTERISTICS FOR EACH REGIME. THE BASELINE REGION SPECTROGRAM IS CHARACTERIZED BY MORE NOISE AND SPORADIC EPISODES OF HIGH SPECTRAL ENERGY FROM NORMAL MOVEMENT AS THE SUBJECT MOVES FROM POINT TO POINT IN THE FACILITY OR BUMPS INTO THINGS. THE COLD REGION SPECTROGRAM SHOWS REDUCED SPECTRAL COMPONENTS, PRIMARILY DUE TO THE DAMPENED MOVEMENTS IN WATER. WE SEE THE SLIGHT BUT IDENTIFIABLE 10 HZ COMPONENT DUE TO THE START OF LIGHT SHIVERING IN THIS REGION. THE COLDEST REGION SPECTROGRAM SHOWS MUCH MORE PRONOUNCED SPECTRAL ACTIVITY, PARTICULARLY IN THE 10 HZ REGION AND THE 10 HZ HARMONICS DUE TO SHIVERING. WE CAN USE

THESE SIMPLE ACCELEROMETER SPECTRAL FEATURES TO BE ABLE TO DISCRIMINATE BETWEEN THESE REGIONS. .... 132

FIGURE 27: LIVE.NET SYSTEM AND ASSOCIATED SENSOR NETWORK USED FOR THE MGH DEPRESSION STUDY. IN THIS STUDY, WE ARE MEASURING BODY MOTION, SKIN CONDUCTANCE (ON THE FINGERS AND BACK OF THE ARM), HEART RATE, HEAT FLUX/TEMPERATURE, AND VOICE ACTIVITY TO FIND HOW THESE TYPES OF NON-INVASIVE PHYSIOLOGY AND DERIVED BEHAVIORAL MEASURES CORRELATE TO DEPRESSION. THE PATIENTS WEAR THE SYSTEM CONTINUOUSLY FOR THE ENTIRE DURATION OF THEIR STAY (TYPICALLY RANGING FROM TWO WEEKS ONWARD) FOR ELECTRO-CONVULSIVE THERAPY AT THE MGH IN-PATIENT PSYCHIATRIC WARD. .... 148

FIGURE 28: SELF-REPORTED EMOTION RATING DATA FOR SUBJECT 6 OVER A 3-WEEK TREATMENT PERIOD. EACH SCALE VARIES FROM BETWEEN -100 TO 100 ON THE Y-AXIS, AND TIME PROGRESSES TO THE RIGHT. WE CAN SEE FROM THE RATINGS THAT THE ECT TREATMENT WAS SUCCESSFUL. LOOKING AT THE RELATIVE TRENDS FOR CORRELATED OUTCOMES SUCH AS DEPRESSION, GUILT, FEELINGS OF HOPELESSNESS, AND SADNESS, THEY ALL SIGNIFICANTLY DECREASE APPROXIMATELY MIDWAY THROUGH THE TREATMENT PROCESS (IN WEEK TWO OF THE SUBJECT’S 3 WEEK STAY), WHILE FEELINGS OF SELF-WORTH, OPTIMISM, ENERGY LEVELS, AND SOCIABILITY ALL INCREASE OVER THE SAME PERIOD. .... 155

FIGURE 29: BOX AND WHISKER PLOT OF DAILY-AVERAGED EMOTION RATING DATA FOR SUBJECT 6. BOXES HAVE LINES AT THE LOWER QUARTILE, MEDIAN (RED), AND UPPER QUARTILE VALUES, AND WHISKERS SHOW EXTENT OF THE REST OF THE DATA. OUTLIERS ARE SHOWN AS ‘+’S BEYOND THE ENDS OF THE WHISKERS. WE CAN SEE THAT FROM THE DIAGRAM THAT THERE IS A HIGH DEGREE OF VARIATION IN MANY OF THE EMOTION RATINGS OF INTEREST, SPECIFICALLY THOSE RELATED TO DEPRESSION..... 156

FIGURE 30: PARETO DIAGRAM OF THE PCA DECOMPOSITION FOR THE RATINGS DATA FOR SUBJECT 6. FROM THE DIAGRAM, WE SEE THAT THE FIRST THREE PRINCIPAL COMPONENTS EXPLAIN 46%, 16%, AND 10% OF THE VARIANCE RESPECTIVELY, OR A TOTAL OF 72% OF THE OVERALL VARIANCE. .... 157

FIGURE 31: LINE CHARTS SHOWING THE RELATIVE COMPOSITIONS OF THE FIRST THREE PRINCIPAL COMPONENTS FOR THE EMOTION DATA FOR SUBJECT 6. PC COMPONENT 1 CAN BE INTERPRETED AS VALENCE, PC COMPONENT 2 AS EMOTIONALITY, AND PC COMPONENT 3 AS ANXIETY. .... 159

FIGURE 32: PARETO DIAGRAM OF THE VALENCE AND EMOTIONALITY EMOTION RATINGS FOR SUBJECT 6. THE DIAGRAM SHOW THAT THE VARIATION IN THE DATA PRIMARILY STEMS FROM THE FIRST TWO COMPONENTS, WITH 76% AND 13% OF THE EXPLAINED VARIANCE RESPECTIVELY. .... 160



FIGURE 33: PCA OF DEPRESSION RATINGS DATA SHOWING THREE CLUSTERS OF NEGATIVE, POSITIVE, AND STRESS EMOTIONS. BLUE VECTORS INDICATE CONTRIBUTIONS OF THE EMOTIONS TO THE FIRST TWO PRINCIPAL COMPONENTS, RED DOTS INDICATE THE DATA POINTS OF THE DIFFERENT DAYS DURING TREATMENT (ALL DAYS TO THE LEFT OF THE Y-AXIS ARE FROM THE FIRST HALF OF THE TREATMENT STAY, AND ALL DAYS TO THE RIGHT ARE FROM THE SECOND HALF). ..... 160

FIGURE 34: CROSS CORRELATIONS FOR  $p < 0.05$  (LEFT) AND RAW VALUES (RIGHT) BETWEEN EMOTION RATINGS FOR SUBJECT 6. AS WE CAN SEE, MANY OF THE EMOTIONS SUCH AS ‘DEPRESSED’, ‘STRESSED’, ‘GUILTY’, ‘EMOTIONAL’, ‘HOPELESSNESS’, AND ‘SAD’ ARE HIGHLY CORRELATED AND ANTI-CORRELATED TO FEELINGS OF ‘SELF-WORTH’, ‘OPTIMISTIC’, ‘ENERGY LEVEL’, ‘EXCITED’, ‘SOCIABLE’, AND ‘HAPPY’. ..... 162

FIGURE 35: PLOT OF NORMALIZED DAILY EMOTION INDICES FOR SUBJECT 6. NOTE THE INVERSE RELATIONSHIP BETWEEN THE NEGATIVE AND POSITIVE EMOTION INDICES. THE COMPOSITE EMOTION INDEX COMBINES THE POSITIVE EMOTION INDEX WITH THE INVERSE OF THE NEGATIVE EMOTION INDEX TO CREATE A COMPOSITE MEASURE WITH THE SAME TRENDING AS THE POSITIVE EMOTION INDEX. ... 163

FIGURE 36: CORRELATION MAPS OF RAW PHYSIOLOGY FEATURES AND EMOTIONAL RATINGS OVER A 5-MINUTE AND 1 DAY WINDOW SIZES. THE BLUE BOX SHOWS CORRELATIONS BETWEEN THE PHYSIOLOGY FEATURES, GREEN BOX SHOWS CORRELATIONS BETWEEN THE EMOTIONAL RATINGS. THE RED RECTANGLE SHOWS THE CROSS-CORRELATIONS BETWEEN THE PHYSIOLOGY FEATURES WITH THE EMOTIONAL RATINGS. THESE CORRELATION MAPS DEMONSTRATE THAT THE CROSS CORRELATIONS BETWEEN THE PHYSIOLOGY FEATURES AND EMOTION RATINGS BECOME STRONGER OVER LONGER WINDOW SIZES. .... 164

FIGURE 37: AGGREGATED PHYSIOLOGY (LEFT) AND AUDIO (RIGHT) FEATURES AND EMOTION RATING CORRELATIONS FOR SUBJECT 6 AT THE GRANULARITY OF A DAY. NOTICE THAT THESE AGGREGATED CORRELATIONS ARE MUCH MORE HIGHLY PRONOUNCED THAN IN FIGURE 36. WE USE STEP-WISE REGRESSION TECHNIQUES TO FIND THE MOST CORRELATED FEATURES TO USE FOR REGRESSION ANALYSIS WITH THE SUBJECTIVE AND CLINICAL OUTCOMES. .... 165

FIGURE 38: CROSS-CORRELATIONS BETWEEN THE VOICE FEATURES. THE FEATURES ARE: AVERAGE LENGTH OF VOICED SEGMENT (ALVS), AVERAGE LENGTH OF SPEAKING SEGMENT (ALSS), FRACTION OF SPEAKING TIME (FST), VOICING RATE (VR), MEAN/STANDARD DEVIATION OF THE FUNDAMENTAL FORMANT FREQUENCY (MFF/SFF), MEAN/STANDARD DEVIATION OF THE CONFIDENCE IN THE FUNDAMENTAL FORMANT FREQUENCY (MCFF/SCFF), MEAN/STANDARD DEVIATION OF THE SPECTRAL ENTROPY (MSE/SSE), MEAN/STANDARD DEVIATION OF THE VALUE OF LARGEST AUTOCORRELATION

PEAK (MVACP/SVACP), MEAN/STANDARD DEVIATION OF THE LOCATION OF LARGEST AUTOCORRELATION PEAK (MLACP/SLACP), MEAN/STANDARD DEVIATION OF THE NUMBER OF AUTOCORRELATION PEAKS (MNACP/SNACP), MEAN/STANDARD DEVIATION OF THE ENERGY IN FRAME (MEF/SEF), MEAN/STANDARD DEVIATION OF THE TIME DERIVATIVE OF ENERGY IN FRAME (MTEF/STEF), CONVERSATION LEVEL (CL), STRESS LEVEL (SL). ..... 169

FIGURE 39: PARETO DIAGRAMS OF PCA DECOMPOSITION OF AUDIO AND PHYSIOLOGY FEATURES. WE SEE THAT THE FIRST PRINCIPAL COMPONENT IN BOTH CASES EXPLAINS THE MAJORITY OF THE VARIANCE OF THE DATA, AND THUS IS A GOOD CANDIDATE FOR A POTENTIAL FEATURE TO USE IN PREDICTING THE EMOTION RATINGS. .... 170

FIGURE 40: WEEKLY CLINICAL OUTCOMES OVER 3-WEEK PERIOD FOR SUBJECT 6 (FOR CGI, CGISEV WAS USED). WE SEE THAT THE ECT THERAPY HAS SUCCESSFULLY TREATED THE SUBJECT, REDUCING A SEVERE CLINICAL DEPRESSION TO SOMETHING THAT IS BELOW A MILD DEPRESSION AS RATED BY THE HAM-D. CORRESPONDINGLY, THE CGI SEVERITY SCORE DECREASES AND THE QLS INCREASES OVER THE SAME PERIOD. .... 171

FIGURE 41: CORRELATION MATRIX OF CLINICAL OUTCOMES AND SELF-REPORTED EMOTION INDICES (PSD = PATIENT SELF-REPORTED DEPRESSION, -EF = NEGATIVE EMOTION INDEX, +EF = POSITIVE EMOTION INDEX, CEF = COMPOSITE EMOTION INDEX). ALL CORRELATIONS ARE IN THE RANGE BETWEEN 0.82-0.99. .... 172

FIGURE 42: SUBJECTIVE COMPOSITE EMOTION INDEX AND HAM-D, CGI, AND QLS CLINICAL OUTCOMES FOR SUBJECT 3. SUBJECT 3 WAS ONLY ON THE WARD FOR 11 DAYS, AND SHOWED STEADY IMPROVEMENT FROM THE ECT THERAPY AS SHOWN IN BOTH THE SUBJECTIVE AND CLINICAL OUTCOMES. SHE WAS RELEASED TO AN OUTPATIENT TREATMENT PROGRAM AT THE END OF THE SECOND WEEK. .... 177

FIGURE 43: HAM-D, CGI, AND QLS CLINICAL OUTCOMES FOR SUBJECT 4. SUBJECT 4 STAYED ON THE WARD FOR OVER 8 WEEKS, AND INITIALLY SHOWED IMPROVEMENT FROM ECT THERAPY TREATMENT THROUGH THE SECOND WEEK, BUT PROGRESS THEN STALLED UNTIL AFTER THE FIFTH WEEK ..... 179

FIGURE 44: SUBJECTIVE COMPOSITE EMOTION INDEX AND HAM-D, CGI, AND QLS CLINICAL OUTCOMES FOR SUBJECT 5. SUBJECT 5 STAYED ON THE WARD FOR ONLY ONE AND A HALF WEEKS, BUT SHOWED IMPROVEMENT TO ECT THERAPY TREATMENT THROUGHOUT HIS STAY ON THE WARD. .... 180

FIGURE 45: A CONDITIONED HIDDEN MARKOV MODEL FOR HIGH-LEVEL BEHAVIOR MODELING. INPUTS FOR  
CONDITIONING ON HOUR OF DAY (H) AND WEEKDAY/WEEKEND (W) AS WELL AS BEHAVIORAL INPUTS.  
WE EXPLORE THE EFFECTS OF VARIOUS INPUTS SUCH AS BEHAVIORAL MEASURES, PHYSIOLOGY  
FEATURES, AND TIME ON THE ACCURACIES OF THE MODELS IN PREDICTING THE HIGH-LEVEL BEHAVIOR  
OF AN INDIVIDUAL ..... 193

# List of Tables

TABLE 1: ACTIVITY CONTEXT CLASSIFICATION ACCURACIES. USING A COMBINATION OF SPECTRAL FEATURES AND POSTURE/ORIENTATION FEATURES USING A GAUSSIAN MIXTURE MODEL, WE CAN ACHIEVE ACTIVITY RECOGNITION ACCURACIES BETWEEN 81.3-100%. ..... 110

TABLE 2: TOP FEATURES CORRELATED WITH BLUFFING. THE HEART RATE VARIABILITY LF/HF RATIO, FRACTION OF SPEAKING TIME, AND MOTION ENERGY WERE THE FEATURES MOST HIGHLY CORRELATED TO BLUFFING..... 119

TABLE 3: TOP FEATURES CORRELATED WITH ALL-IN PLAYS. THE NUMBER OF SKIN CONDUCTANCE PEAKS, THE STD. DEV. OF THE RR INTERVALS OF THE HEART, AND VOICE PITCH VARIATION AS MEASURED BY THE STD. DEVIATION OF THE FORMANT FREQUENCY WERE THE MOST CORRELATED FEATURES..... 121

TABLE 4: TOP FEATURES CORRELATED WITH BAD-BEAT PLAYS. VOICE ENERGY, MOTION ENERGY, AND VOICE PITCH VARIATION AS MEASURED BY STD. DEVIATION OF THE FUNDAMENTAL FORMANT FREQUENCY WERE THE MOST CORRELATED FEATURES. .... 122

TABLE 5: TOP FEATURES CORRELATED WITH STRESS. THE NUMBER OF SKIN CONDUCTANCE PEAKS, THE STD. DEV. IN THE LOCATION OF THE LARGEST AUTO CORRELATION PEAK, AND THE LF/HF RATIO HEART RATE VARIABILITY MEASURE WERE THE MOST CORRELATED FEATURES..... 122

TABLE 6: CLASSIFIER RESULTS FOR STRESS AND BLUFFING. WE CAN PREDICT HIGH STRESS EVENTS (DEFINED AS A SUBJECTIVE RATING OF 7 FOR STRESS ON A 10-POINT SCALE) WITH 82% ACCURACIES AND BLUFFING (DEFINED WHEN THE SUBJECT INDICATED A LARGE SCALE BLUFF) TO WITHIN 71% ACCURACIES. .... 123

TABLE 7: TOP FEATURES SELECTED FOR MULTIVARIATE REGRESSION AGAINST COMPOSITE EMOTION INDEX (CHOSEN WITH A  $p < 0.05$  CRITERION) FOR SUBJECT 6. STD. DEV. OF THE LOCATION OF THE LARGEST AUTOCORRELATION PEAK, MAXIMUM VALUE IN ACCELERATION, AND MAXIMUM IN THE SUM OF ACCUMULATED DIFFERENCES WERE THE TOP THREE PERFORMING FEATURES. .... 166

TABLE 8: CHART SHOWING MULTILINEAR REGRESSION USING THE TOP TWO FEATURES (STD. DEV. OF THE VALUE OF THE LARGEST AUTOCORRELATION PEAK IN THE AUDIO, AND MAX. VALUE OF THE ACCELEROMETER), ACROSS THE 26 DIMENSIONS OF SELF-REPORTED EMOTIONS AS WELL AS FOR THE DEPRESSION EMOTION INDICES FOR SUBJECT 6. THE  $R^2$  OF ALL THE EMOTION INDICES ARE BETWEEN 0.79-0.85..... 168

TABLE 9: CLINICAL OUTCOME REGRESSION RESULTS FOR SUBJECT 6 USING TOP TWO FEATURES CHOSEN IN STEPWISE REGRESSION ANALYSIS FOR THE COMPOSITE EMOTION INDEX, ALONG WITH THEIR BETAS. 173

TABLE 10: BEHAVIORAL FEATURES CORRELATIONS WITH COMPOSITE EMOTION INDEX FOR SUBJECT 6. SEDENTARY BEHAVIOR WAS THE BEST BEHAVIORAL PREDICTOR FOR THE SUBJECT, FOLLOWED BY THE IN-BEDROOM MEASURE AS DETERMINED BY THE LOCATION BEACONS. .... 175

TABLE 11: TOP FEATURES SELECTED FOR MULTIVARIATE REGRESSION AGAINST COMPOSITE EMOTION INDEX (CHOSEN WITH A  $p < 0.05$  CRITERION) FOR SUBJECT 3. STD. DEV. OF THE VALUE OF THE LARGEST AUTOCORRELATION PEAK, STD. DEV. IN MOTION, STD. DEV. IN LOCATION OF THE AUTOCORRELATION PEAK WERE THE TOP THREE PERFORMING FEATURES. .... 177

TABLE 12: SUBJECTIVE AND CLINICAL OUTCOME REGRESSION RESULTS FOR SUBJECT 3 ALONG WITH THEIR BETAS. ALL CORRELATIONS FOR A TWO FEATURE MODEL HAVE AN  $R^2$  BETWEEN 0.55-0.87. .... 178

TABLE 13: TOP FEATURES SELECTED FOR MULTIVARIATE REGRESSION AGAINST COMPOSITE EMOTION INDEX (CHOSEN WITH A  $p < 0.05$  CRITERION) FOR SUBJECT 4. STD. DEV. OF FUNDAMENTAL FORMANT, MOTION ENERGY, AND MOTION ENTROPY WERE THE TOP THREE PERFORMING FEATURES. .... 179

TABLE 14: SUBJECTIVE AND CLINICAL OUTCOME REGRESSION RESULTS FOR SUBJECT 4. ALL CORRELATIONS FOR A TWO FEATURE MODEL HAVE AN  $R^2$  BETWEEN 0.47-0.89. .... 180

TABLE 15: TOP FEATURES SELECTED FOR MULTIVARIATE REGRESSION AGAINST COMPOSITE EMOTION INDEX (CHOSEN WITH A  $p < 0.05$  CRITERION) FOR SUBJECT 5. STD. DEV. OF THE VALUE OF THE LARGEST AUTOCORRELATION PEAK, STD. DEV. IN MOTION, MAXIMUM NUMBER OF PEAKS IN MOTION WERE THE TOP THREE PERFORMING FEATURES. .... 181

TABLE 16: SUBJECTIVE AND CLINICAL OUTCOME REGRESSION RESULTS FOR SUBJECT 5. ALL CORRELATIONS FOR A TWO FEATURE MODEL HAVE AN  $R^2$  BETWEEN 0.72-0.95. .... 181

TABLE 17: TOP FEATURE SETS PERFORMANCE ACROSS SUBJECTS USING THE STD. DEV. OF THE LOCATION OF THE AUTOCORRELATION PEAK AND THE STD. DEV. IN THE RAW MOTION FEATURES. HAM-D OUTCOMES CAN BE CORRELATED WITH AN  $R^2$  BETWEEN 0.57-0.84, CGI WITH AN  $R^2$  BETWEEN 0.77-0.95, AND QLS WITH AN  $R^2$  BETWEEN 0.43-0.73. .... 182

TABLE 18: LIFESCAPE BEHAVIOR MODELING RESULTS USING A CONDITIONED THREE-STATE HMM  
CORRESPONDING TO HOME/WORK/OTHER ACTIVITY CLASSES WITH A VARIETY OF INPUT FEATURE  
COMBINATIONS. THE BEST PERFORMING MODEL HAS TIME OF DAY INPUTS AS WELL AS AND BOTH  
LONGITUDINAL AND TRANSVERSE ACCELEROMETER CHANNELS. ....195

# Chapter 1 Introduction

Over the upcoming decades, dramatic changes in healthcare systems are needed worldwide. In the United States alone, 76 million baby boomers are reaching retirement age within the next decade [Tavormina 1999]. Current healthcare systems are not structured to adequately service the growing needs of the aging population, and a major crisis is imminent. The current system is dominated by infrequent and expensive patient visits to physician offices and emergency rooms for treatment of illness. The failure to do more frequent and regular health monitoring is particularly problematic for the elderly with multiple co-morbidities and often tenuous, rapidly changing health states. Even more troubling is the fact that medical specialists cannot explain how most problems develop because they usually only see patients when things have already gone wrong.

Given this impending healthcare crisis, it is imperative to extend healthcare services from hospitals into home environments. Although there has been little success in extending health care into the home, there clearly is a huge demand. In 1997, Americans spent \$27 billion on health care outside of the health care establishment, and that amount has been increasing [Abowd 2003]. Moreover, our dramatically aging population makes it absolutely necessary to develop systems that keep people out of hospitals. By 2030, nearly one out of every two households will include someone who needs help performing basic activities of daily living and labor-intensive interventions will become impractical because of personnel shortage and cost [Abowd 2003].

The best solution to these problems lies in proactive healthcare technologies that put more information and control into the hands of the people, such as the Guardian Angel System [Szolovits et al. 1994]. Out-of-pocket spending by US consumers for alternative therapies has begun to rival that for traditional health services [Eisenberg et al. 1998]. Personalized health systems will enable a trend toward decentralizing healthcare away from resource-intensive hospitals. The vision is a healthcare system that will help an individual maintain their normal health profile by providing better monitoring and feedback, so that the earliest signs of disease can be detected and corrected. This can be accomplished by continuously monitoring a wide range of vital signs, providing early warning systems for people with high-risk medical problems, and “elder care” monitoring systems that will help keep seniors out of nursing homes.

Most available commercial mobile healthcare platforms have focused on data acquisition applications, with little attention paid to enabling real-time, context-aware applications. Companies such as VivoMetrics, Bodymedia, and Mini-Mitter, have extended the basic concept of the ambulatory Holter monitor (enabling a physician to record a patient's ECG continuously for 24-48 hours), which for three decades has been the only home health monitor with widespread use [Holter 1949].

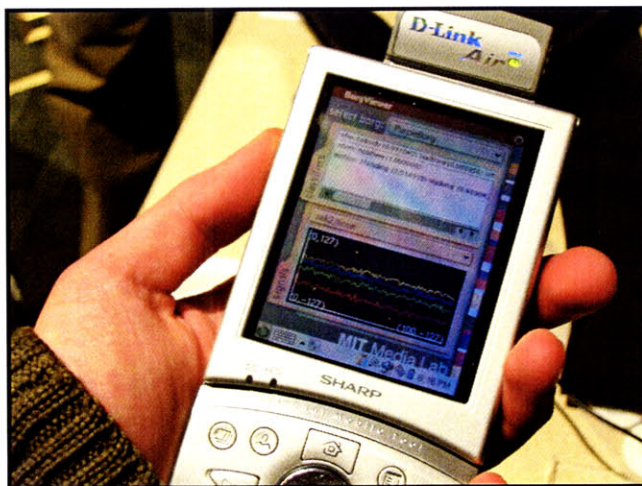
Additions to this individual monitoring paradigm have been extended along two fronts: medical telemetry and real-time critical health monitoring. Regarding the former, various in-patient medical telemetry systems have been developed in recent years, focusing on providing an infrastructure for transporting and storing data from the patient to caregivers for later analysis [Bashshur et al. 1997]. In terms of the latter, a few systems have extended the health monitoring concept by augmenting a physiological monitor (usually based on a single physiological sensor) with specialized algorithms for real-time monitoring within specific application domains, such as heart arrhythmia, epileptic seizures, and sleep apnea, which can potentially trigger alerts when certain critical conditions or events occur [FDA 2003, Kirk et al. 2003]. However, the development of proactive healthcare technologies beyond these basic telemedicine and individual event monitoring applications has been rather slow. The main limitation has been the high costs and inflexibility of limited monitoring modalities associated with these technologies and the impracticality for long-term use in general settings.

Enter the current mobile device revolution. In 2004, over 600 million mobile phones were sold, dwarfing the number of personal computers sold that year. It can definitely be argued that the first truly mass-marketed wearable devices were mobile phones and personal data assistants (PDAs). While the primary purpose of these devices is for point-to-point voice communications (mobile phones) or personal data organization (PDAs), we are seeing increasing convergence in the functionality of these devices as general purpose computing machines, with flexible data communications capabilities. In fact, the computing power in today's smart phones and other mobile convergence devices are now reaching those of personal desktops systems in the mid 1990s. Likewise, 3G technologies provide large bandwidth channels to transmit and share data in a real-time fashion.



Yet despite the advances in mobile technologies, modern day mobile devices are still limited in a fundamental way. The devices are effectively senseless devices which have no notion of a user's context except what is explicitly provided to it by the user. The missing element in modern day mobile technology is integrated sensing technologies that can allow devices to monitor a user's context and act proactively on the user's behalf by providing useful feedback.

Mobile devices make an ideal platform for long-term monitoring of individuals as people habitually carry the devices with them at all times. Thus, if mobile devices are empowered with sensing capabilities, there would be an opportunity for these devices to identify the immediate context of the individual: what the user is doing, where the user is, who the user is interacting with, what the person's immediate health state is, etc. Furthermore, it would be possible to learn patterns in the user's behavior and context, since a mobile device is continually in direct contact with a user over long-periods of time. This potentially allows these devices to make inferences regarding the future state of the user for proactive applications.



**Figure 1: LiveNet wearable performing real-time FFT analysis and activity classification on accelerometer data, visualizing the results, and wirelessly streaming real-time ECG/skin conductance/temperature and classification results to a remote computer with a projection display as well as peer LiveNet systems (which may be located anywhere in the world).**

In this thesis, we make progress toward integrating flexible sensing, communications, and real-time processing infrastructure with existing mobile device technology to enable practical context-aware applications. Specifically, we have developed a flexible distributed mobile platform called LiveNet that has been deployed for a variety of proactive healthcare applications.

Based on cost-effective commodity mobile device hardware with customized sensors and data acquisition hub plus a lightweight software infrastructure, LiveNet is capable of local sensing, real-time processing, and distributed data streaming. This integrated monitoring system can also leverage off-body resources for wireless infrastructure, long-term data logging and storage, visualization/display, complex sensing, and computation-intensive processing.

The LiveNet system allows people to receive real-time feedback about their continuously monitored and analyzed health state. In addition, LiveNet can communicate health information to caregivers and other members of an individual's social network for support and interaction. Thus, by combining general-purpose commodity hardware with specialized health/context sensing within a networked environment, it is possible to build a multi-functional mobile healthcare device that is at the same time a personal real-time health monitor, multimodal feedback interface, context-aware agent, and social network support enabler and communicator.

A number of key attributes and capabilities of the LiveNet system that make it an enabling distributed healthcare system include:

- Modular architecture based on commodity hardware that can be improved with time
- Wireless capability with resource posting/discovery and data streaming to distributed endpoints
- Flexible sensing for context-aware applications that can facilitate interaction in a meaningful manner and provide relevant and timely feedback/information
- Unobtrusive, minimally invasive, and non-distracting sensing technology
- Continuous long-term monitoring capable of storing a wide range of physiology as well as contextual information
- Real-time classification/analysis and feedback of data that can promote and enforce compliance with healthy behavior

- Capturing long-term behavioral trends of repeating patterns of behavior and subtle physiological cues that can be used to track and model user behavior as well as to flag deviations from normal baselines
- The capability to create new forms of social interaction and communication for community-based support by peers and establishing stronger social ties within family groups

With the development of increasingly powerful diagnostic sensing technology, doctors can obtain more context-specific information directly, instead of relying on a patient's recollection of past events and symptoms, which tend to be vague, incomplete, and error prone. Although many of these specialized sensing technologies have improved with time, most medical equipment is still a long way off from the vision of cheap, small, mobile, and non-invasive monitors. Modern imaging technology, for example, costs thousands of dollars per scan, requires room-sized equipment chambers, and necessitates uncomfortable and time-consuming procedures.

While there is always a place for more invasive physiological sensing technologies, the medical community often focuses unnecessarily on only the clinical "gold-standards" for research and monitoring. To really push health technologies to the masses, however, it is necessary to make the sensors used by medical devices as minimally invasive and unwieldy as possible. Personal health systems must be lightweight, easy-to-use, unobtrusive, flexible, and non-invasive to make headway as viable devices that people will use. Reducing the burden on people is critical for developing practical systems that people can wear for long periods of time. This is imperative over the next couple decades, as proactive healthcare technologies will need to gain traction to keep people out of hospitals to reduce the strain on the healthcare system.

As such, there is a potentially huge area of exploration to use basic non-invasive sensing technologies as proxies for more-invasive diagnostic sensing devices [Sung & Pentland 2005]. The LiveNet system focuses on using combinations of non-invasive sensing and contextual features (specifically, measures such as heart rate, motion, voice features, skin conductance, temperature/heat flux, and location information) instead of more involved clinical physiology sensing such as pulse oximetry, blood pressure, and multi-lead ECG. Gold-standard sensing technologies typically require expensive and bulky monitoring equipment, careful professional

preparation and calibration of sensors, and more direct interfacing of the sensors with the patient. As a result, a preponderance of health monitoring tasks severely limits the mobility patient and typically cannot take place over long periods of time. In contrast, we will demonstrate that it is possible to enable monitoring and relevant clinical classification applications over long-periods of time using simple, cheap sensors that do not require laborious setup procedures and can be comfortably worn and can allow patients to go about their daily activities unimpeded.

By using these non-invasive sensors, the LiveNet system can continuously monitor physiology, motor activity, sleep patterns, and other indicators of health. The data from these sensors can then be used to build a personalized profile of performance and long-term health over time, tailored to the needs of the patients and their healthcare providers. This unique combination of features also allows for quantification of personal contextual data, such as amount and quality of social interactions and activities of daily living. This type of information is potentially very useful for increasing the predictive power of diagnostic systems. It is through these types of sensing that systems will become more “human-aware” about the world surrounding the user: who they are interacting with, what they are doing, what their health or cognitive state is, and other contextual information.

Essentially, what we can measure determines what we can detect. Many behaviors such as exercise, behavior patterns, activity patterns, gait, and sleep can be potentially quantified using non-invasive sensors such as accelerometers. We can consider these types of behaviors as a class of “new vital signs” that have been demonstrated to exhibit large correlations to many of the important medical conditions (such as arthritis, heart disease/stroke, cancer, diabetes, depression, and dementia) that will have to be addressed by future healthcare systems as populations age [Fauchet et al. 2004].

The research in this thesis demonstrates that it is possible to use simple non-invasive physiological and contextual sensing using the LiveNet system to accurately classify a variety of interesting contextual situations. First, it is demonstrated that non-invasive sensing data can be correlated to a variety of important physiological and behavioral phenomenon, and thus can serve as substitutes to more invasive and unwieldy forms of medical monitoring devices while still providing a high level of diagnostic power. This demonstrates the basic feasibility of using simply obtained measures such as heart rate, motion, skin conductance, and audio features over

other more involved physiological sensing that is traditionally used. This is important in ambulatory and continuous monitoring applications, where more cumbersome sensing such as typically found in medical and clinical research settings is not possible. Special attention is given to develop robust methods of classification given cheap, simple sensors (in contrast to the professional medical-grade equipment that is typically emphasized in clinical classification), even in the presence of noise, motion artifacts, sloppily-placed sensors, and other sub-optimal conditions.

From this foundation, the LiveNet platform is deployed in two studies to classify some interesting contextual states based on non-invasive sensing. The first study demonstrates that it is possible to use simple motion sensing to classify the immediate activity context of an individual. In particular, accurate real-time classifiers are developed using only a single accelerometer sensor capable of identifying a diverse set of activity contexts such as walking, sitting down, working at the office, watching television, etc. The second study attempts to show that it is possible to use non-invasive physiological sensors to identify stress and interest in the structured setting of poker tournaments. We demonstrate that we can even use physiological sensing to identify human intent such as lying.

Another obvious domain for using a platform such as LiveNet is in real-time soldier physiology monitoring and other critical health monitoring applications. For this reason, we initiated a research collaboration with the US Army Research Institute of Environmental Medicine (ARIEM) at Army Natick Labs to demonstrate the feasibility of using the LiveNet system for these types of critical healthcare applications. The ultimate goal is to be able to develop a variety of low-burden, real-time wearable health monitors capable of classifying a variety of conditions that soldiers may exhibit through the course of training or combat. As a pilot study, the LiveNet platform has been used to collect motion and body temperature data of volunteer soldiers who are immersed in cold water for extended periods of time. It is demonstrated that it is possible to develop real-time cold-exposure classifiers based solely on non-invasive motion sensing. We show that the LiveNet system can be used to robustly classify these types of clinically significant physiological conditions in the presence of noise, motion artifacts, and other adverse conditions found in real-world situations.

We also demonstrate that the LiveNet system is relevant in terms of the practical continuous monitoring of long-term physiological trends in normal patterns of human behavior. Continuous monitoring ensures the capture of relevant events and the associated physiology wherever the patient is, expanding the view of healthcare beyond the traditional outpatient and inpatient settings. Long-term monitoring has the potential to help create new models of health behavior. For example, long-term monitoring may provide important insights into the efficacy and effectiveness of medication regimes on the physiology and behavior of a patient over time, at resolutions currently unobtainable.

In a clinical study with the Psychiatry Department at the Massachusetts General Hospital, the LiveNet system is used to correlate physiological sensing to depression state, as well as to track changes in depression state over time through the course of treatment via continuous monitoring. Thus, this thesis shows how long-term physiology changes can be captured to objectively measure medical treatment and medication efficacy. In parallel to this clinical study, the LiveNet system is used to collect data on people's long-term everyday life. The physiological data that is collected is used to conduct long-term modeling and trending of high-level behavior. These long-term trends, which we term a person's "lifescape," are detected based on simple physiologic and contextual information, demonstrating the possibility of predicting high-level human behavior based solely on ambulatory mobile sensing technology.

These various studies show the range and breadth of the LiveNet platform in allowing classification over a wide range of interesting domains, including the physiology of clinical conditions, affective state and the effects of treatment/medication over time, and contextual information such as activity and stress state. Thus, this thesis demonstrates the ability of using the LiveNet platform to develop practical real-time monitoring systems to assist patients manage their day-to-day health state as well as to elucidate personal trends in physiology over time.

## 1.1 Thesis Contributions

This thesis makes four principal contributions:

**Contribution 1: Practical Commodity Hardware-based Wearable Monitoring Systems.** *We have developed and deployed a wearable sensing and monitoring platform based on low-cost commodity hardware that can be used to continuously monitor individuals in long-term ambulatory settings.* In this thesis we show how we can leverage modern-day mobile devices such as smart phones and personal data assistants (PDAs) to develop flexible systems capable of multimodal sensing and monitoring, real-time data streaming, context-aware inference, and distributed networking. The wearable hardware, biosensor, and systems infrastructure that has been developed is capable of supporting a variety of relevant healthcare applications. In addition, we have developed a suite of LiveNet applications that allow us to easily conduct a variety of large-scale data acquisition, monitoring, and real-time feedback applications. We show that LiveNet can be used as a practical vehicle to conduct clinically significant research in both laboratory and ambulatory settings. We have demonstrated that the LiveNet platform can be used to support clinically significant research, specifically in the context of health state classification for such medical conditions as cold exposure/hypothermia and stress. We also demonstrate the potential of using the LiveNet system for long-term continuous monitoring in clinical settings, specifically to assess the efficacy of treatment/medication on severely depressed individuals.

**Contribution 2: Clinical and Behavioral Datasets.** *We have generated a number of novel datasets exploring human physiology and behavior, spanning a number of interesting domains.* Using the LiveNet platform, we have been able to quantitatively data-mine long-term human physiology and behavior. The ARIEM Cold-Exposure Dataset contains complete movement, heart rate, heat flux, and gold-standard core body temperature data for over thirty volunteer U.S. Army Ranger subjects as they are immersed in cold-water environments. The PokerMetrics Dataset includes physiology, behavior, interest, and outcome data of amateur poker players in thirty heads-up poker tournaments. The MGH Depression Dataset currently includes continuous long-term physiology and ambulatory behavior of six clinically-depressed patients receiving electroconvulsive therapy at the MGH psychiatric ward, for over a cumulative fifteen weeks of continuous 24-hour monitoring (this study is ongoing). The MGH depression study, in particular, is the first of its scope to the author's knowledge. The Lifescape Dataset consists of two and a

half years (over 22,000 continuous hours) of long-term physiological/behavioral data on three subjects going about their daily lives.

**Contribution 3: Non-Invasive Context-Aware Sensing.** *We demonstrate how non-invasive sensing can be used to create effective context-aware systems.* We show that simple, minimally-invasive, noisy, non-expensive physiologic sensing features can be correlated to, and used as, effective substitutes to more invasive health sensing for robust context classification, thereby enabling practical mobile health applications in long-term continuous ambulatory settings. We have also developed a few general contextual classifiers for identifying interesting contexts such as activity, stress level, cold-exposure, and depression state based on simple physiologic cues that can be detected using non-invasive sensing.

**Contribution 4: Health Trending and Behavior Modeling.** *We use the data collected to uncover how continuously-monitored physiology and immediate contextual state can be correlated to long-term health state and behavior.* We demonstrate that long-term monitoring can allow one to track trends in human physiology in a quantitative, continuous manner that hasn't been possible in the past given technological limitations. Using the obtained physiological and behavioral data, we have developed accurate models that can trend a patient's subjective emotions and clinical outcomes through the course of medication and treatment. We also show that we use non-invasive physiology data to develop generative probabilistic graphical models that can be used to classify and predict an individual's long-term behavior and time-dependent dynamics.

## 1.2 Thesis Roadmap

The content of this thesis is grouped into ten principal chapters, as follows:

**Chapter 1 Introduction:** A brief overview of the motivations and overall themes of the thesis research is provided. We introduce the LiveNet platform, and how the system can be used to study human behavior and physiology and enable new proactive applications that extend existing healthcare paradigms. We then discuss the main contributions of this thesis and briefly outline the content in the subsequent chapters.



**Chapter 2 Background:** We first provide the history of wearable healthcare and telemedicine technologies, as relevant in the domains of the healthcare and biomedical industries. We then comment on the available spectrum of health technologies, deconstructed in the context of features that serves to differentiate the basic attributes of the technologies in a meaningful way. For each technology feature, the historical context up until the present is given first. The status quo is then contrasted to methodologies that extend the existing paradigm in a more general and useful manner, and then the ways in which the LiveNet platform that was developed in this thesis can potentially be used to enable such extensions is discussed.

**Chapter 3 Wearable and Sensing Infrastructure:** In this chapter, we start by detailing the various hardware and sensing technologies that were developed for the LiveNet platform, centered around a sensor hub called the SAK2 board. We then discuss the capabilities of the Enchantment software infrastructure that enables the real-time processing and data streaming functionality of the system. After that, a discussion of the various human factors concerns in developing practical wearable systems for use in long-term ambulatory settings is addressed. We conclude with a description of a number of applications that were developed to facilitate data collection and monitoring, as well as a few demo applications that showcase the flexibility and power of the LiveNet platform. We conclude by discussing a number of research collaborations that have used the LiveNet platform.

**Chapter 4: Non-Invasive Features and Modeling:** This chapter outlines the different types of non-invasive physiological and behavioral sensing that the thesis focuses on, including movement, audio/voice, temperature/heat flux, heart rate, and skin conductance. For each type of sensing, we go in-depth into the physiological significance. We then detail the various types of features that can be extracted to be used in the analysis of the data from the studies that we have conducted. We also describe a number of the modeling and classification techniques that are used in the analysis of the data collected from our studies. We conclude by discussing how low-cost non-invasive sensing can correlate to more gold-standard clinical physiology sensing.

**Chapter 5 Non-Invasive Context Classification:** In this chapter, we demonstrate how the non-invasive sensing that was described in the previous section can be applied to develop interesting context-aware systems. In particular, we show how motion sensing enables us to develop very powerful classifiers capable of differentiating between a variety of everyday

activities. The second half of the chapter focuses on the PokerMetrics Study, where we use non-invasive physiology to develop accurate classifiers that can be used to detect stress and lying.

**Chapter 6 Real-Time Clinical Health Classification:** This chapter is devoted exclusively to the ARIEM Soldier Monitoring Study that was conducted in conjunction with the U.S. Army Natick Labs. We detail first the protocol, then show how we can use motion sensing to develop highly accurate real-time shivering monitors. We then demonstrate how we can use Hidden Markov Modeling techniques to develop models of the time dynamics of shivering and correlate motion to core body temperature. The resulting classification systems can triage a soldier's cold exposure state with near-perfect accuracy.

**Chapter 7 Continuous Clinical Monitoring and Trending:** This chapter details the MGH Depression Study, which monitors the continuous physiology of clinically depressed patients at the psychiatric ward at the Massachusetts General Hospital. We detail the protocol, and then follow up with an in-depth analysis of how the physiology sensing and behavior data can be correlated to the subjective emotion ratings of the patients. We then follow by showing how these features can be used to create multi-linear regression models that can accurately trend clinical outcomes of depression.

**Chapter 8 Long-Term Behavior Modeling:** This chapter describes the Lifescape Study, where over a year and-a-half of long-term physiology and behavioral data was collected on a number of subjects. We demonstrate how we can use conditioned hidden Markov model as a means of behavior parameterization and prediction. We model the long-term behavioral dynamics of individuals going about their daily lives solely with the use of non-invasive physiological sensing.

**Chapter 9 Future Directions:** We discuss potential future directions in which to extend the existing corpus of research that this thesis demonstrates. Specifically, we discuss how our research can be extended to develop real-time feedback systems enabling rich proactive healthcare applications in the domain of clinical diagnosis and eldercare. We also describe how the existing LiveNet platform can be used to develop distributed healthcare systems that allow for social support.

**Chapter 10 Conclusions:** The thesis concludes with a discussion of the major accomplishments of the thesis, and the current direction this technology may be taking society. We discuss potential ramifications of how long-term continuous monitoring systems will help reshape our understanding of human behavior, and the implications in a number of academic disciplines.

## Chapter 2 Background

There has been extensive literature written on how new technologies will affect the future landscape of healthcare [Pentland-A 2004, Wilson 1999]. In the following chapter, we provide background on how new technologies are being integrated into the healthcare industry. We first provide the history of wearable and telemetry technologies as relevant in the domains of the healthcare and biomedical industries. We then comment on the current spectrum of health technologies, deconstructed along dimensions that serve to differentiate the basic attributes of the technologies in a meaningful way. For each technology dimension, the historical context up until the present is given first. The status quo is then contrasted to new methodologies that extend the existing paradigm in a more general and useful manner, and how infrastructure such as the LiveNet platform that was developed in this thesis can potentially be used to enable such extensions.

### 2.1 History of Wearable Health Technology

The concept of using wearable technology as a method for improving an individual's senses and capabilities is centuries old. Many consider the first written record of a wearable device to be attributed to Roger Bacon, who in 1268 made the earliest recorded comment on the use of lenses for eyeglasses. Four hundred years later, Robert Hooke, in the preface of his famous work *Micrographia*, proposed the possibility of a variety of devices that could improve human sensing and functioning, much as eyeglasses had improved sight, stating:

“By the addition of such *artificial Instruments* and *methods*, there may be, in some manner, a reparation made...[by] rectifying the operations of the *Sense*, the *Memory*, and *Reason*...in respect of the Senses, is a supplying of their infirmities with Instruments, and as it were, the adding of artificial Organs to the natural...And as Glasses have highly promoted our seeing...there may be found many mechanical inventions to improve our other senses of hearing, smelling, tasting, and touching.”

These insights were well before their time, and did not find application until hundreds of years later, when devices such as the electric hearing aid [c. 1899] and implantable artificial pacemaker

[c. 1950] finally began to appear as practical medical devices. Around the same time that the artificial pacemaker was invented, the first medical monitoring device was developed by a physician named Norman Holter, who in 1949 invented an ECG (electrocardiogram) device eponymously named the Holter monitor. This device was a 75-pound backpack that could record the ECG signals from the wearer and transmit the signals [Holter 1949]. The monitor was quickly reduced in size in subsequent years (using magnetic tape and eventually electric recording) and developed into the first widespread ambulatory ECG device. More recently, smaller portable 'event recorders' have become widespread for patients with heart conditions such as intermittent arrhythmias. Instead of continuously monitoring over a period of time, a patient places the monitor to his chest to record an ECG signal when he perceives an intermittent palpitation [FDA 2003].

Unfortunately, these Holter monitors and variants have remained the only real practical ambulatory medical monitoring devices well into the dawn of the second millennium. In recent years, there have finally been substantial advances in portable biomedical monitoring technology, as a variety of wearable medical monitors have been developed for the space research/military and are starting to become commercially viable both in hospital and home-use settings.

Military and space exploration have always been interested in soldier and astronaut monitoring applications and have driven early mobile health monitoring applications. A joint project between Georgia Institute of Technology and the U.S. Navy called the Georgia Tech Wearable Motherboard (GTWM) was started in mid 1990s [Gopalsamy et al. 1999]. The GTWM is a wearable vest that is woven with optical fibers that transmit signals until interrupted, such as by a bullet wound to a soldier. Peripheral sensors to measure heart rate, temperature and respiration can be incorporated through its flexible bus architecture. In 2002, a joint collaboration between Stanford University and NASA called Lifeguard [Mundt 2002] formed to design a health monitor for eventual use on astronauts for remote physiological monitoring. The wearable system component, called the CPOD, is a Sony Walkman-sized unit worn around the torso and connects to ECG leads, respiration sensor, a pulse oximeter, three-axis accelerometer, temperature probe, and a potential add-on blood pressure monitor. Data can be transmitted wirelessly to a Bluetooth or RF base station.

One of the first areas of commercialization in health monitoring technology was in the exercise and fitness markets. Devices such as heart rates monitors from market-leader Polar [Polar] or FitSense [FitSense] are now ubiquitous. FitSense has even augmented their traditional line of heart rate monitors with mini-wearable sensors that can communicate health information to a monitoring device, including pedometers, core body temperature sensors, and location tags, which have been used for marine and army ranger training and as demonstration implementations for the WPMS program (Warfighter Physiological Monitoring System) of the U.S. Army.

Just in the last few years, examples of portable health monitoring technology commercialization have finally moved into the home, and include companies such as Digital Angel [Digital Angel], which is developing a watch with wireless connection, GPS, and a biosensor array (pulse, ECG, temperature, blood pressure, pulse oximeter) for the elder care market. If a patient travels outside a pre-set boundary, falls and is down for at least one minute, encounters a significant change in ambient temperature, or triggers the emergency button, an alert is sent wirelessly to a caregiver. Another company called BodyMedia [BodyMedia] has developed an ergonomically designed armband that continuously measures activity (2-axis accelerometer), galvanic skin response, near body temperature, and skin temperature, heat flux and heart rate. BodyMedia also recently created a strategic alliance with Roche Diagnostics to use their sensing technology to create a system for weight maintenance.

Within the professional physiological monitoring space, VivoMetrics has created a wearable vest called the LifeShirt, which is embedded with inductive plethysmographic sensors, a two-axis accelerometer, and a single-channel ECG that can continuously monitor an ambulatory patient. Additional peripheral devices can be optionally added to record blood pressure, blood oxygen saturation and, end tidal CO<sub>2</sub> can be plugged into the recorder [FDA 2001]. MiniMitter's VitalSense [MiniMitter] is an integrated physiological monitoring system, with core body temperature, dermal temperature, and heart rate sensing.

Others groups have focused on the miniaturization of sensing systems. The Berkeley Smart Dust Project has used its Motes in health applications to create small ECG systems and other health monitors [Malan et al. 2004]. The Wearable Computing Lab at ETH Zurich has developed a wrist monitoring device called AMON (Advanced telemedical MONitor) that measures vital signs (pulse, oxygen saturation, blood pressure, ECG, skin temperature) [Lukowicz et al. 2002]. A ring-

worn wearable monitor, developed by the D'Arbeloff Labs, contains a plethysmographic sensor and RF transceiver in a compact form-factor that can continuously monitor vital signs such as pulse oximetry, heart rate, blood pressure, and saturated oxygen levels [Asada et al. 2003].

## 2.2 History of Telemedicine

The invention of the telegraph in 1835 and telephone in 1876 provided the infrastructure necessary for medical telemetry applications. Even prior to the invention of television, people already envisioned the use of audiovisual technology as a means for doctors to care for patients remotely, as illustrated on the cover of Radio News in April 1924. In 1905, Willem Einthoven, the pioneer of electrocardiography, was the first to transmit a medical signal via a telephone wire when he sent an ECG signal between his lab and a hospital 1.5 km away, creating the first 'telediagram' [Reiser 1978]. From there, a number of telemedicine projects sprouted up: the first transmission of radiological films over telephone lines took place in between West Chester and Philadelphia, Pennsylvania in 1948; radiologists at Jean-Talon Hospital in Montreal, Canada created a teleradiology system in the 1950s; and the University of Nebraska developed a two-way teleconferencing system to conduct neurological examinations across campus in 1959; and the first use of wireless telecommunications for physical diagnosis was demonstrated by Dr. Kenneth Bird in 1968, using a microwave audiovisual link between a medical station at Logan Airport in Boston and the Massachusetts General Hospital [Murphy & Bird 1974].

NASA played a pioneering role in the early development of telemedicine and provided much of the technology and funding for early telemedicine demonstrations in the early 1960s, when humans began flying in space during the Apollo missions. Physiological parameters were transmitted via radio from both the spacecraft and the space suits during missions. These early efforts and the enhancement in communications satellites fostered the development of telemedicine applications [Bashshur et al. 1975]. The STARPAHC project, a joint effort by the U.S. Indian Health Services, NASA, and Lockheed Company provided health care to the isolated Papago Indian tribe and to astronauts [Bashshur 1980].

In modern day use, Holter monitors worn by patients for monitoring cardiac rhythms can initiate telemetric transfers to specialists. In 2001, the FDA approved the first wearable defibrillator [FDA 2003] that also connects to an external modem for transmission to a specialist. The

Micropaq from Welch Allyn [Micropaq] measures ECG signals, heart rate, and SpO<sub>2</sub>, and sends this data and potentially patient-triggered alarms via wireless LAN to the nurses' station for monitoring.

## **2.3 Gold-Standard vs. Non-invasive Sensing**

As early as 400 BC, the Greeks had already begun transforming medicine from superstitious craft into a science. The earliest reference of modern diagnostic sensing is attributed to Hippocrates, known as the founder of modern scientific medicine, who noted that audible sounds emanated from the heart. William Harvey revolutionized the theory of medicine by noting the heart's role in blood circulation, and correlated heart sounds to the pumping of the heart. However, it was French doctor Rene Laennec who invented the first medical diagnostic sensing device in 1816, the stethoscope (combining the Greek words for 'I see' and 'the chest'). His focus was on auscultation, or the diagnosis of sounds from the body, and he invented the device after being embarrassed at needing to put his ear on the chest of a female patient.

From these humble beginnings, modern medicine has been revolutionized by diagnostic devices that could provide observable measurements of physiology and pathology which would then be objectively analyzed by medical professionals. Prior to the institutionalization of diagnostic sensing techniques, doctors had to make diagnoses based mainly on verbal discussions with the patient. With the development of increasingly powerful diagnostic sensing technology, doctors could obtain more diagnostic information directly, instead of relying on a patient's recollection of past events and symptoms, which could be vague, incomplete, and prone to error.

Amazingly enough, it turns out that a combination of a few basic physiological measures have huge amounts of diagnostic power, and are often adequate on their own to provide an accurate medical diagnosis. It is for this reason that these basic physiologic measures are the things the doctor always checks first in a medical examination, most notably, ECG/heart rate and heart morphology information, SpO<sub>2</sub>/blood oxygen, blood pressure, respiration, and temperature.

Beyond these basic measures, modern sensing these days includes a huge variety of diagnostic tools that are very powerful for specialized cases, including electromyography (EMG) or muscle activity, electroencephalography (EEG), blood testing, x-rays, ultrasound, to CAT (computed



axial tomography), PET (positron emission tomography), MRI (magnetic resonance imaging), etc., that all play very important diagnostic and monitoring roles.

While many of these specialized sensing technologies have improved with time, most medical equipment is still a long way off from the ideal of cheap, small, mobile, and non-invasive monitors. An MRI or CAT scan today can still cost thousands of dollars per scan, require room-sized equipment chambers, and necessitate uncomfortable and time-consuming procedures. In fact, patient compliance issues even exist for ambulatory ECG Holter monitors despite the critical nature of heart conditions, precisely because of the invasive nature of electrode leads.

As such, there is a potentially huge area of exploration in the use of basic non-invasive sensing as proxies for more-invasive diagnostic sensing devices. It is a goal of this thesis to demonstrate that a combination of non-invasive sensing features (specifically heart rate, motion, audio features, skin conductance, and temperature) can be correlated with physiology that typically requires more specialized, invasive sensing such as pulse oximetry, blood pressure, and ECG. Here, we extend the target to not just traditional medical diagnostic data but also contextual data such as social interaction, activities of daily living, etc., as this data can contain extremely useful contextual information with predictive power.

In this thesis, we demonstrate that a combination of non-invasive sensing features (specifically heart rate, motion, voice features, skin conductance, and heat flux/temperature) can complement, and sometimes replace, physiology that typically requires more specialized, invasive sensing such as pulse oximetry, blood pressure, and multi-lead ECG. Here, the goal is not just to suggest alternatives to traditional medical diagnostic data but also to use this type of sensing to determine user context in quantifying things such as social interaction, activities of daily living, etc. It is through these types of sensing that systems will become more “human-aware” about the world surrounding the user: who they are interacting with, what they are doing, where they are, what their health or cognitive state is, and other contextual information.

Specifically, we demonstrate that it is possible to develop robust methods of classification given cheap, simple sensors (in contrast to the professional medical-grade equipment that is typically emphasized to do clinical classification), even in the presence of noise, motion artifacts, sloppily-placed sensors, and other sub-optimal conditions. We show how we can use non-invasive

measures to derive features within a statistical machine learning and modeling framework to accurately classify a variety of clinical physiological states, disease pathologies, affective states, and even human behavior.

## **2.4 Post Processing vs. Real-time Feedback**

If one looks at the adoption of technology for wearable research since the field came into being in the early 1990s, one would find people focused on customized electronics solutions such as the MIThril Project at MIT [MIThril] or the Spot Project at Carnegie Mellon [Spot] because of the lack of appealing options found in industry. Of paramount importance was the fact that processing power, storage, form factor, flexibility of I/O options, and display modalities could simply not be found in existing commercial systems [Starner 2001], and so these groups created their own.

However, in just the last few years, mobile technology advances have gotten to the point where a person can cheaply afford to buy a device with significant computational power and the I/O capabilities of fully general computers in small form factor. This has created an inflection point as mobile technologies start to be able to support really practical applications that can do localized data analyses and classification on streaming data, with interesting output modalities and networking options.

Prior to this point, the dominant healthcare paradigm that existed was to stream physiology data from an individual to a centralized server, where the higher-power processing and data visualization could be performed to post-process the data. In such a system, a wearable system served mainly as a data acquisition vehicle, with little feedback or interaction capabilities. Now, it is possible for significant localized processing as well as displaying the result, which will open up the door to real-time interactive health applications. In fact, commercial systems are just beginning to incorporate these types of increased functionality. The Welch-Allyn Micropaq represents an advancement in telemedicine systems, as it is one of the only available systems that also provides a local display on the device that can display basic medical information.

An obvious domain for the LiveNet platform, then, is in physiology monitoring with real-time feedback and classification. In this thesis, we demonstrate that mobile systems are capable of significant local processing for real-time feature extraction and context classification as well as

provide the distributed wireless infrastructure for streaming information between systems. This enables the real-time classification of medical conditions without the need for other infrastructure, available wherever the individual goes. The distributed nature of LiveNet can also allow systems to stream raw physiology or its combination with derived metadata/context very easily to any specified source(s), whether it is other mobile systems, data servers, or output displays such as projectors.

By providing local processing capabilities with ambulatory sensing, the time that is required to receive feedback for relevant health events is dramatically reduced. Historically, the time delay required to receive feedback can potentially take weeks, and be problematic because of the iterative nature of determining the optimized treatment path. For people who are on medication or embarking on a prolonged rehabilitation schedule, for example, this delay in the feedback loop is particularly onerous. This same scenario is also true with lengthy rehabilitation programs such as cardiac rehabilitation. A doctor will recommend a dosage and medication regimen to try out, and the person goes home and tries the medication schedule out for a while. If the person does not respond favorably to this drug schedule, they have to reschedule an appointment with the doctor, go in, and potentially take more tests, before getting a recommendation on a new schedule. This results in a very time-consuming process as well as a significant drain on healthcare resources. In addition to the fact that the feedback loop can be very long in duration, the doctor is in the dark about the efficacy of the treatment, and so an iterative trial-and-error process is required.

By using an ambulatory system with real-time feedback, we can develop more personalized medication and rehabilitation scheduling based on measured, quantitative physiological symptoms and behavioral responses, not based only on time scheduled approximations as the current practice. The effects of medication and rehabilitation treatment can be logged and recorded quantitatively and compared to changes in physiology, eventually with the goal of developing a real-time, closed-loop monitoring and drug delivery system that can track the effects of individualized treatment over time. An example of such research is a study that has demonstrated that one can track the effects of Parkinson's medication to treat the symptoms of tremoring using only non-invasive accelerometry [Weaver 2003].

Thus, with real-time sensing and processing, it is possible to effectively reduce the time delay in processing and receiving health feedback. This is particularly true when the doctor can be

removed from the equation, such as with real-time diagnostic systems that can provide effectively instantaneous classification on health state and context. Medication compliance, for example, is a major healthcare issue (especially among the elderly) with estimated costs of upwards of \$100 billion annually [Werheimer 2001] and can be enforced using a real-time alert system to help people comply with healthy, preventative behavior.

The research in this thesis shows that we can develop practical real-time systems that can track the effects physiology or treatment over time. In particular, the ARIEM Cold-Exposure Study in Chapter 6 demonstrates the ability for LiveNet to implement real-time classifiers to monitor the health state of soldiers in the field. Likewise, the MGH Depression Study described in Chapter 7 shows how it is possible to enable individuals and caregivers to know how their physiology changes to treatment or medication instantly, and closes the loop on the delay that exists during adjustment of conditions to suit the needs of the patient.

## **2.5 Event Reporting vs. Continuous Monitoring**

The first continuous health monitor was the abovementioned Holter monitor, invented by Dr. Norman Holter in 1949 for continuously monitoring of the heart for up to 24 hours [Holter 1949]. Prior to ambulatory devices that could continuously monitor physiology, patients could only report relevant events (e.g. “I had horrible chest pain”), with no other information available about these events that a doctor could use to help with diagnosis.

Continuous monitoring, on the other hand, can empower doctors or clinical trial investigators to view physiologies during specific events, whether adverse or otherwise relevant (not just snapshots). The essential characteristic is that the continual monitoring ensures the capture of relevant events and the associated physiology wherever the patient is, not just at the hospital. In addition, long-term monitoring can provide insight on the effects and efficacy of medication regimes on the physiology of a patient over time, at resolutions currently unobtainable.

Minimally invasive sensing expands the relevance of monitoring systems outside the clinical context in terms of the potential to being able to monitor everyday behavior. As technology shrinks and becomes less and less unwieldy and invasive, it is feasible that we are on the horizon of a world where it may be commonplace for people to wear long-term health monitoring systems

even when they are not sick. The point is to move to a more proactive healthcare paradigm in which computer-assisted technology tracks personal health trends and non-healthy behaviors are monitored and corrected before they become chronic (that on top of the standard critical health monitoring applications that people focus on today).

LiveNet promises to be especially effective for continuously monitoring medical treatments. Currently, doctors prescribe medications based on population averages rather than individual characteristics, and they check the appropriateness of the medication levels only occasionally—and expensively. With such a data-poor system, it is not surprising that medication doses are frequently over- or underestimated and that unforeseen drug interactions occur. Stratifying the population into phenotypes using genetic typing can improve the problem, but only to a degree.

Continuous monitoring of motor activity, metabolism, and so on can be extremely effective in tailoring medications to the individual. For patients to function at their best, their medications must be optimally adjusted to the diurnal variation of these symptoms. For this to occur, the managing clinician must have an accurate picture of how a patient's symptoms fluctuate throughout a typical day's activities. In these situations, a patient's subjective self-reports are not typically very accurate, so objective clinical assessments are necessary. Continuous monitoring provides the means for a physician to objectively track clinical outcomes.

Time-stamping events, in order to correlate these events with accurate continuous physiology and behavior data in an ambulatory setting, also offers great potential. From this health information, real-time correlations to specific medical conditions as well as predictions of adverse outcomes can be made. This has a potentially large impact in the research of physiology in the domain of clinical medication and intervention research. For example, continuous monitoring providing a streamlined path for Electronic Data Capture (EDC), where accurate reporting is an issue and human transcription errors and recall bias from surveys can be reduced. Potential benefits provided by continuous monitoring include automated and real-time data capture from patients for accurate reporting, feedback and notification for enforcing medication compliance in patients, assessment of the degree of medication compliance, removal of human error inherent in manual transcription/data entry, high-resolution time-stamping for accurate temporal characterization of events, critical physiological monitoring for clinical trial drug safety protocols in patients and the

ability to accurately correlate quantitative physiologic data to events for diagnosis and characterization.

In the MGH Depression Study described in Chapter 7, we show how we can use continuous physiology and behavioral monitoring to track treatment and depression state over time in clinically depressed individuals.

## **2.6 Short-term vs. Long-term Monitoring**

Almost all existing health monitoring systems past and present, with the possible exception of BodyMedia, were developed without the intention of monitoring physiology and health state for long periods of time, on the order of weeks to months. In the past, systems were designed for instantaneous detection of certain events, or are used over a period of time (on the order of hours to days) in order to capture certain events that are infrequent. With the advent of high-density storage media and the miniaturization of electronics, it is now possible to start capturing the long-term physiology of one's life. In fact, physiology information is very low-bandwidth in contrast to other media formats such as audio/video. It is estimated that a few gigabytes of storage is ample to store a person's multimodal physiology for their entire life, and this is already in reach of current miniaturized compact flash/SD memory cards.

It turns out that medicine has not focused on the science of physiological changes over the long term, primarily because in the past there was no way of monitoring continuously throughout a person's life. Instead, long-term physiological monitoring currently consists of spot checks from physical exams, which occur on the order of months if not years. Thus, it is only within critical health domains that people have explicit notions of their semi-continuous physiology, such as diabetics who need to know their blood sugar levels to properly regulate their health maintenance.

This mentality will change with time because of technology improvements that will make an individual's own health state more explicit. This is imperative for more proactive health management paradigms. Currently, people do not have any technologies that can help them monitor and track their health state, thus relying on hospitals when they become sick to fix a problem that perhaps could have been fixed had more health information been available. With

improved self-awareness of their own physiology, they can be more proactive to maintain their health before they become sick and have to go to the hospital.

While the LiveNet system has been designed with comfort and ease of use in mind, it is noted that it may not yet be practical and burdenless for the general population to wear PDA-based systems for long periods of time. However, it is important to point out that the LiveNet system is agnostic to the underlying hardware, and within a few generations (i.e., a couple of years), new much smaller, faster, integrated-functionally cellphones which a person could conceivably slip into their pocket and forget about will exist. The LiveNet system therefore demonstrates a system that is flexible enough to eventually evolve into something that will be appropriate for practical future long-term health monitoring applications.

A great deal of progress in terms of understanding human physiology will result from the fact that long-term trends in human physiology can be explored in detail. Such advances include tracking the development and evolution of diseases, monitoring changes in physiology as people grow older, comparing physiology across different populations (gender, ethnicity, etc), and even knowing characteristic physiology patterns of people who are *healthy* (this last example is particularly important when it is necessary as a diagnostic methodology designed to quantitatively define abnormal behavior). From continuous monitoring, a very fine granularity of quantitative data can be obtained, in contrast to the surveys and history-taking that has been the mainstay of long-term studies and health interventions to date.

The goal is to be able to detect repeating patterns in complex human behavior by analyzing the patterns in data collected from the LiveNet system. In this thesis, we demonstrate that the LiveNet system can be used for long-term, continuous monitoring to gain insights into patterns in long-term physiology that have not been explored in medical literature. Specifically, in Chapter 7 we describe how we can track the long-term depression state of clinically depressed patients.

## **2.7 Reactive vs. Proactive Technologies**

Technology is progressing such that one is just beginning to see real-time health monitoring systems, which can determine if certain medical conditions arise, typically in the domain of critical monitoring applications such as such as heart arrhythmia, epileptic seizures, and sleep apnea detection [FDA 2003, Kirk et al. 2003]. These systems are purely reactive (e.g. sounding

an alarm after a person has a heart attack or falls down), and are dependent on classifying and determining in real-time when certain events have occurred.

While reactive systems are very useful and potentially-life saving, it is also a limiting concept. It is very important from the proactive healthcare point of view that future health systems also be able to determine trends in physiological/contextual state over time, to provide not only immediate diagnostic power but also prognostic insight. By combining long-term trending with multimodal analysis, it is possible to develop more proactive systems that can catch problems before they manifest themselves. Thus, instead of just being able to monitor and react to a critical health condition in real-time, systems can start to look at long-term patterns of physiology and begin to predict a person's future health state.

It is commonly estimated that up to 80% of medical problems may result from counterproductive behavior. Thus, a lot of focus in proactive health monitoring systems [Intille 2003] is on how to elicit change toward alternative healthy actions at the moment of decision, which is hypothesized to be where individuals can be persuaded to change their behavior, dubbed by Steven Intille as 'Persuasive Computing'. Similarly, Vadim Gerasimov developed bio-analytical games using small biomonitoring wearable systems [Gerasimov 2003] to evaluate how an interactive game can assist in personal medical management.

However, a proactive system requires more resources, as it must have context-aware and inference capabilities to be able to determine what the right information is to be directed at the right people, to the right places, at the right times, and for the right reasons. This is an additional level of burden, and small advances have been made in this regard. Context-aware annotation prompting systems have been implemented in the past to reduce user load. An example of utilizing context-awareness is experience sampling, a technique to gather information on daily activity by point of querying (which can be set to trigger based on movement or other sensed context by the PDA), or even relational devices which attempt to query "nicely" so as to minimize annoyance [Intille 2002]. A platform such as LiveNet is particularly amenable for developing these types of healthcare applications, especially those requiring context-awareness with local real-time processing and localized interaction modalities [Sung & Pentland 2004, Sung & Pentland 2005, Sung-B et al. 2005].



A perfect example of an application domain in which proactive technology can help is in maintenance of medication cycles. Since there is usually a delay before the effects of medication take hold and wear off, knowing when to dose a patient is partly trial and error. Usually, methodologies simply dose at regular intervals without heed for how the dosage might affect the patient, which may vary according to diet, exercise, and a myriad of other factors. As such, there is typically an efficacy cycle when taking medication, with the effects of medication undershooting or overshooting the ideal target most of the time. However, a system that can monitor the appropriate long-term physiology or behavioral outcomes (which may be idiosyncratic to an individual) can proactively determine when to actively maintain dosage so the desired effects of medication can be constant.

The LiveNet platform provides a convenient infrastructure to implement and rapidly prototype new proactive healthcare applications in this domain. The Lifescape Study in Chapter 8 shows that it is possible to use long-term contextual and non-invasive physiological monitoring to accurately trend and model long-term behavior. These models, in turn, can then be used to flag deviations from the predictions from normal behavior. Thus, the trending and modeling research from this study shows how the LiveNet system may be capable of supporting more proactive, intelligent healthcare applications than currently exists.

## **2.8 Pervasive Infrastructure vs. Mobile Technology**

There has been a dichotomy in the mindsets of researchers as to the best methodology for implementing future healthcare applications. One philosophy focuses on putting sensing and technology infrastructure in the environment, freeing the individual from having to carry anything. Pervasive computing, ambient intelligence, and ‘smart rooms’ all fit into this paradigm (such as the House\_n/TIAX PlaceLab Consortia [PlaceLab] and the Rochester Center for Future Health Smart Home [URCFH]). While this methodology is attractive as it makes the technology transparent, it has potential downsides in terms of large implementation costs, the hard problems with analysis (such as with machine vision), as well as the fact that the technology is limited to being effective only where infrastructure exists. Furthermore, it is often not technically feasible or practical to obtain some types of information from sensing at a distance (such as physiological information). Although promising headway has been made in remote at-a-distance physiological

sensing, such as Nexsense's technology (which may one day allow heart monitors or MRIs using Doppler-based technology [Nexense], there are still many problems with noise, transmission artifacts, and other issues that need to be overcome before this becomes a reality, and the preponderance of physiological signals will require local contact for the foreseeable future.

At the other end of the spectrum, the wearable computing community has focused on developing mobile health technology that people can carry at all times, wherever they go. While this provides the opportunity to be able to constantly and intimately monitor all the relevant contextual information of an individual as well as provide instantaneous feedback and interaction, it comes at the cost of potentially being invasive, unwieldy, and distracting.

There exist rich opportunities that lie somewhere in between these two polar methodologies. By combining elements from each type of technology, it is possible to leverage the benefits of both to create a system that is more flexible than either alone. LiveNet attempts to do this, creating a powerful mobile system capable of significant local sensing, real-time processing, distributed data streaming, and interaction, while relying on off-body resources for wireless infrastructure, long-term data logging and storage, visualization/display, complex sensing, and more computation-intensive processing.

## **2.9 Centralized vs. Distributed Computing**

Originally, for hundreds of years, the healthcare system was a very decentralized system where doctors would make personalized house calls to treat the ailments of patients. This distributed healthcare model persisted until the advent of specialized diagnostic technology, which played a significant role in the centralization of healthcare services. As modern diagnostic equipment and medicine began to be very specialized, expensive, and non-portable, health and medical services began to converge into common centers that could share overhead and more efficiently utilize resources, forming the hospitals which are still the dominant paradigm for modern healthcare systems that one sees in many parts of the world today.

The influence of this paradigm is seen in the prevailing trend in health monitoring and telemedicine applications, which is focused on collecting and sending data uni-directionally to a specialist for analysis and diagnosis, rather than providing local feedback for the patient that is

necessary for more proactive healthcare applications. By having all health information flowing to the doctor instead of empowering the individual, the onus of care is placed on the doctor, resulting in situations where individuals do not even know much about their own health state.

Ironically, even though advancements in medical sensing result in ever increasing diagnostic health information, the patient has simultaneously become increasingly cut out from the loop. However, new advances in ambulatory medical monitoring and context aware systems such as those provided by the LiveNet system will allow healthcare systems to come full circle, allowing patients to stay at home without the inefficiency and expense of going to a hospital. As discussed in the introduction, this is very vital given the stress placed on the healthcare system as a whole.

There is also the possibility of extending the notion of sending health information, not only to a single endpoint, but also to a network of potential caregivers, which may be physically distant (consider specialists in different countries working together and visualizing health information from a remote patient). The LiveNet system is based on a flexible distributed architecture that supports such data streaming to arbitrary endpoints (including one to many, many to one, and many to many). Another possibility for extension is also providing real-time processing and metadata based on the continuous data streams, allowing the local system to be completely self-contained and providing relevant context-aware interaction (potentially determining if an alarm needs to be triggered to caregivers, or initiating a reminder or suggestion for proactive behavior modification). Whereas most telemedicine systems provide only raw physiological data to be streamed to a remote monitoring station, systems such as LiveNet that contain a local mobile device can have significant processing capabilities to do complex real-time classification tasks as well as to provide real-time data visualization, information feedback, and multimodal interaction modalities.

Because the LiveNet system is a distributed system that is capable of streaming data to arbitrary endpoints, one can imagine a variety of telemedicine applications where real-time feedback of important health state can be relayed to both individuals as well as to doctors and other caregivers. By having a distributed system, health information can be distributed not only one way to a centralized specialist, but also put back in the hands of people who should be most invested in their health regulation — the patients themselves. A distributed architecture allows for a more distributed healthcare system, functioning at multiple levels, including for local self-regulation

and feedback to the patient, to support group/family/peers, as well as to other specialists who may be in different locations.

## 2.10 Individualized vs. Group Applications

Most health applications have focused on proactive feedback to empower individuals with health information, making them more active in their own disease and health management [Wysocki et al. 2003]. More recently, however, people have begun looking at group applications that may be synergistic toward individual health maintenance. Vikram Kumar, in his DiaBetNet research work, established how computer games for young diabetic children can engender competitive play that would as a by-product cause children to more effectively monitor their own blood-sugar concentrations, resulting in better insulin management [Kumar 2004].

This concept can be elaborated into developing general social support networks using the LiveNet system. This form of '*peer-to-peer pressure*' can be defined as using groups of peers in your support/family group to elicit healthy behavior from peer-pressure using a peer-to-peer distributed system to disseminate information. The concept of persuasive computing can be extended to include community support at moment of decision to elicit healthy behavior. Arguably, social peer pressure is one of the strongest forms of persuasion. This is the reason why people form social support groups, and this can be applied in a positive way whether it is to quit smoking, drinking, drugs, etc. Thus, the movement of health information can be extended beyond the critical health monitoring domain for such things as soldier systems and hospital networks, to include the sharing of long-term health and physiology information with peer groups via an explicitly laid out social support infrastructure.

The social sharing of physiology and context-based information can have profound effects in establishing stronger social ties within groups. By providing context information, including audio/video, physiological, and other contextual cues, it is possible to create a sense of virtual proximity between physically separated people. Also known as telepresence, this implicit awareness and sense of closeness to peers is a result of both the reduced communication barriers as well as the fact that meta-information of the other person's health state is available at all times. A similar concept where traditional speech communications is augmented with the user's situational context has been demonstrated (e.g. busy or not to take a phone call, etc.) [Marmasse

2004]. This also has an analog in traditional computer supported collaborative work, where the concept of 'presence' as applied to a user set context, or even by activity state, can be used as an important contextual cue as to whether one should initiate collaborative work [Fogarty 2004].

Thus, one can imagine normal communication that is augmented with non-verbal physiologic and other contextual cues that can trigger responses from the social network in new forms of socio-biofeedback. An open question is how to frame the physiological and contextual information. The LiveNet system provides a convenient platform to explore how technology can assist social interaction, by providing non-distracting, relevant feedback/information, and context-awareness to facilitate interaction in a meaningful manner.

## Chapter 3      **Wearable and Sensing Infrastructure**

It has long been the collective vision of the medical community and technologists alike to find ways in which wireless and mobile technology can enhance the healthcare system. Reality, however, has often failed to live up to these high expectations, as widespread technology adoption in the healthcare industry has historically been very slow. Typically, technology is showcased as demonstration systems in various experimental healthcare technology initiatives, but unfortunately remains as such, without penetration in generalized real-world settings. In addition, many emerging healthcare technologies are often limited to content delivery onto mobile devices, which miss the rich potential for more interactive and context-aware paradigms.

In such situations, issues such as usability, flexibility, and extensibility are often afterthoughts overshadowed by demonstrations of the features of the raw technology. The problem lies in the difficulty of effectively implementing applications in the base technologies without developing the foundation infrastructure necessary to make those technologies effective. These issues must be addressed before deployment of new healthcare technologies can be ubiquitous.

Several key components are necessary to create usable technology infrastructure, which can be effectively applied in healthcare settings. First and foremost, a flexible and scalable system architecture platform is required to appropriately adapt to various healthcare settings, potentially involving up to hundreds of individual users. In addition, the technology must be robust in real-world environments, a requirement that has been the major obstacle of mobile/wearable technologies. Furthermore, relevant technology must have the sensing and computational resources to enable real-time, context-aware interaction [Brown et al. 1997, Starner et al. 1998]. Finally, the human factors side of the equation that has plagued mobile technologies must be properly balanced, and the interface must be appropriately tailored to the application.

In this thesis, we attempt to achieve many of the above lofty goals that have been lacking in existing wearable systems. Toward this end, we have developed a flexible wearable platform called LiveNet intended for long-term ambulatory health monitoring with real-time data streaming and context classification [Sung-B et al. 2005]. The platform is based on the MIThril 2003 architecture [DeVaul et al. 2003], a proven accessible architecture that combines inexpensive commodity hardware, a flexible sensor/peripheral interconnection bus, and a

powerful light-weight distributed sensing, classification, and inter-process communications software layer to facilitate the development of distributed real-time multimodal and context-aware applications. There are three major components to the MITHril LiveNet architecture; a PDA-centric mobile wearable platform, the software network and resource discovery API, and a real-time machine learning inference infrastructure.

The defining feature of the original MITHril project was the modular, distributed, clothing-integrated design based on a unified power/data bus, allowing one to put sensing, computing, and interaction resources where they were most useful and appropriate [DeVaul et al. 2001, Lukowicz 2001]. The MITHril platform was redefined in 2003 with the advent of inexpensive wirelessly enabled PDA hardware. PDAs and other modern mobile devices provide the perfect base platform for multi-user, wireless, distributed wearable computing environment, supporting dozens of interacting users and large-scale interaction and sensing experiments. It is this maturation of commercial technology that is paving the road to viable large-scale healthcare paradigms. Although personal data assistant devices (PDAs) are typically used for individual-user mobile applications, there exists a greater opportunity to tie these PDAs together with a uniform data communication and resource discovery infrastructure that can result in much richer forms of interaction dynamics.

LiveNet demonstrates the ability to use this standardized PDA hardware tied together with a flexible software architecture and modularized sensing infrastructure to create a system platform where sophisticated distributed healthcare applications can be developed. The platform can be used to develop systems that are able to continuously monitor a wide range of physiological signals together with a user's activity and context, to develop a personalized, data-rich health profile of a user over time. The LiveNet system has the following features:

**Hierarchical, distributed modular framework:** The LiveNet platform was designed to be highly modular in order to accommodate the variety of potential applications of interest. For example, we can plug-and-play any number of sensors of interest to individually configure each individual LiveNet system. In addition, all the hardware components such as the particular mobile device used can be replaced with future generation devices with little effort. Given the software infrastructure, we can easily set up distributed applications involving multiple users across a wireless network.

**Based on low-cost commodity PDA/embedded hardware:** Instead of custom-designed embedded systems which have been the mainstay of wearable computing systems in the past, we show that we can leverage low-cost commodity hardware to create powerful context-aware systems. A fully configured LiveNet system composed of an embedded mobile device capable of real-time interaction and distributed data streaming along with the SAK2 sensor hub BioSense board (described in Section 3.2 and Section 3.3, respectively) can be constructed in moderate volumes for under \$150. While the current system implementations are based on PDAs, the software infrastructure is made to be portable to a variety of mobile devices, including cell phones, tablet PCs, and other convergence devices. As such, our systems leverage commercial off-the-shelf components with standardized base-layer communication protocols such as TCP/IP; this allows for the rapid adoption and deployment of these systems into real-world settings.

**Flexibility in accommodating new custom and commercial sensors:** Through the sensor hub of the LiveNet system, it is possible to interface with a wide range of commercially available sensors, including pulse oximetry, respiration, blood pressure, EEG, blood sugar, humidity, core temperature, heat flux, and CO<sub>2</sub> sensors. Any number of these sensors can be combined through junctions to create a diversified on-body sensor network. The LiveNet system can also be outfitted with BlueTooth, Secure Data (SD), or Compact Flash (CF) based sensors and peripherals, and other I/O and communication devices including GSM/ GPRS/ CDMA/ 1xRTT modems, GPS units, image and video cameras, memory storage, and even full-VGA head-mounted displays.

**Real-time data streaming and resource allocation/discovery:** The LiveNet hardware and software infrastructure provides a flexible and easy way to gather heterogeneous streams of information, perform real-time processing and data mining on this information, and return classification results and statistics. This information can result in more effective, context-aware and interactive applications within healthcare settings. The software infrastructure even provides for SSL encryption to allow for secure channels when dealing with sensitive information.

**Allows for rapid application prototyping:** Given the modular nature of the hardware components, abstracted network communications API, established software feature extraction/modeling/inference modules, and flexible interface design tools, we can easily prototype new



applications. In fact, we have demonstrated the creation of instant messaging applications and even VOIP-style communication programs in a matter of days.

In this chapter, we first describe the hardware system infrastructure for the LiveNet system in Section 3.1. The sensing subsystem and sensing technologies are then described in Section 3.2 and Section 3.3, respectively. After this brief technical introduction, some human factors considerations when dealing with mobile applications are then discussed in Section 3.4. Section 3.5 then gives a brief description of the software infrastructure that allows us to provide the data streaming, information resource discovery and posting, and real-time processing capabilities. We then describe some LiveNet applications that were developed using the infrastructure in Section 3.6, and finish with some system rollouts and studies that have utilized the LiveNet platform in Section 3.7.

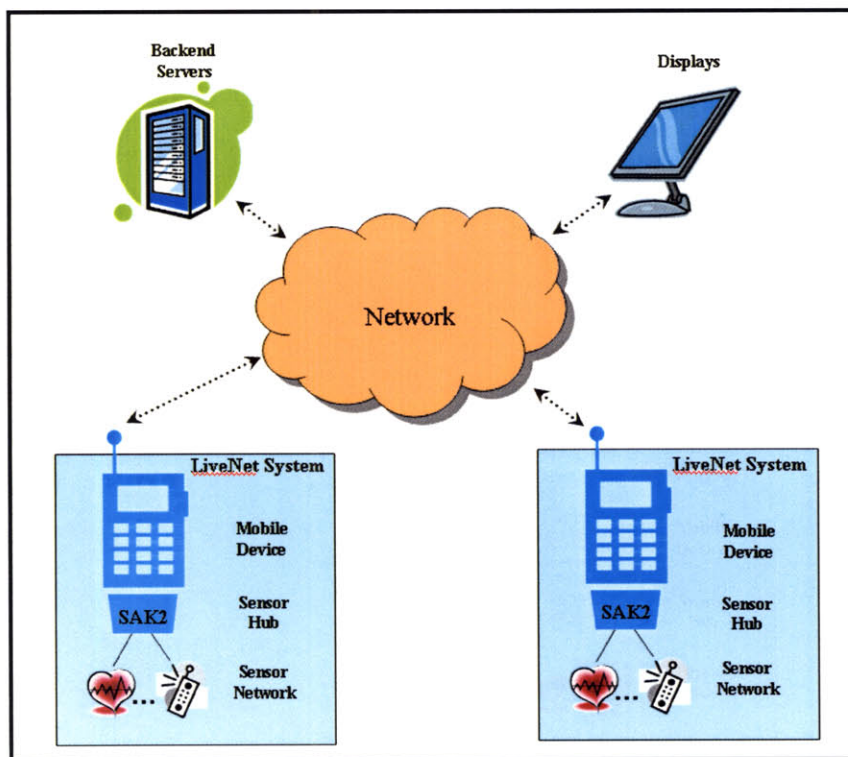


Figure 2: Hierarchical LiveNet Architecture. Each LiveNet system is comprised of a mobile device (with local processing, storage, and wireless capabilities), the SAK2 sensor hub, and a configurable sensor network. The LiveNet architecture can be configured for both traditional client-server applications as well as peer-to-peer communications for distributed applications involving multiple systems. Data streams can be subscribed to by backend servers (for more intensive processing that cannot be handled locally), displays, and other endpoints.

### 3.1 System Hardware Technology

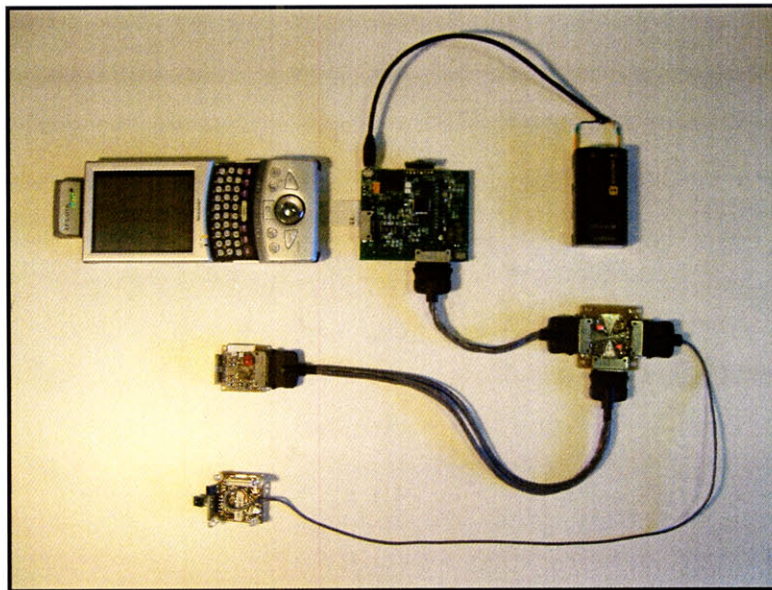
The LiveNet system is currently based on the Sharp Zaurus, a Linux-based PDA mobile device that leverages commercial development and an active code developer community. Although LiveNet can utilize a variety of Linux-based devices, the Sharp Zaurus PDA provides a very convenient platform. This device allows applications requiring real-time data analysis, peer-to-peer wireless networking, full-duplex audio, local data storage, graphical interaction, and keyboard/touch screen input.

The Sharp Zaurus SL-5500, the main mobile device used by LiveNet currently, has the following features:

- 206 Mhz StrongARM processor (400 MHz XScale for the SL-6000)
- Linux-based embedded OS w/ QT Palmtop Environment, Java
- TFT LCD with front light, 240x320 pixel, 65,536 colors
- Memory 64MB SDRAM, 16MB FLASH ROM
- Touch panel input, built-in QWERTY keyboard
- 1 CF slot, 1 SD card slot (enables 802.11 wireless + storage)
- I/O Port Serial/USB (via docking station port), IR port
- Stereo headphone jack, audio input (mono), buzzer / alarm

A sensor hub, called the SAK2 (described below), is used to interface the PDA with the sensor network. The SAK2 is an extensible and “hackable” system that can be used independently for stand-alone data acquisition applications. A small sample of applications using this sensor hub and previous implementations in stand-alone operation includes real-time critical health monitoring [Sung 2002] and identifying activities of daily living [Bao 2003], and social network monitoring [Choudhury 2003, Choudhury & Pentland 2003].

Currently available stand-alone sensor designs include accelerometers, IR active tag readers (used in conjunction with IR tags that can be used to tag locations or objects), battery monitors, GPS units, microphones, ECG/EMG, galvanic skin response (GSR), and temperature sensors. The sensor hub also allows us to interface with a wide range of commercially available sensors, including pulse oximetry, respiration, blood pressure, EEG, blood sugar, humidity, core temperature, heat flux, and CO<sub>2</sub> sensors. Any number of these sensors can be combined through junctions to create a diversified on-body sensor network.



**Figure 3: LiveNet system, composed of the Zaurus PDA (top left), with SAK2 data acquisition/sensor hub and BioSense physiological sensing board (middle), battery source (top right), sensor bus hub (lower right), 3D accelerometer board (middle left), and WMSAD multisensor board (lower left).**

The LiveNet PDA configuration supports wireless IP networking through the CF interface using the 802.11b wireless protocol. This low-cost wireless networking capability is a crucial enabling feature, allowing us to implement multi-node, distributed wearable applications.

The 802.11b protocol provides sufficient bandwidth for many systems to simultaneously stream full-duplex audio and sensor data, as well as send lower-bandwidth data such as text messaging and state information. With wireless connectivity, data can be streamed to off-body resources for a variety of purposes, including data logging and storage, visualization/display, and computation-

intensive off-board processing. Persistent network servers are also used for directory services and resource discovery.

In order to support long-term health monitoring and activities of daily living applications, a specialized extensible, fully integrated physiological sensing board called the BioSense was developed as a add-on board to the sensor hub. The board incorporates a three dimensional (3D) accelerometer, ECG, EMG, galvanic skin conductance, a serial-to-I2C converter (which can be attached to any 3<sup>rd</sup> party serial-based sensing device), and independent amplifiers for temperature/respiration/other sensors that can be daisy-chained to provide a flexible range of amplification for arbitrary analog input signals.

With the combined physiological sensing board and third-party sensors, a fully outfitted LiveNet system can simultaneously and continuously monitor and record 3D accelerometer, audio, ECG, EMG, galvanic skin response, temperature, respiration, blood oxygen, blood pressure, heat flux, heart rate, IR beacons (which can tag locations, objects, or even people), and environmental activity sensors. The sensor data and real-time classification results from a LiveNet system can also be streamed to off-body servers for subsequent processing, trigger alarms or notify family members and caregivers, or be displayed/processed by other LiveNet systems or computers connected to the data streams for complex real-time interactions.

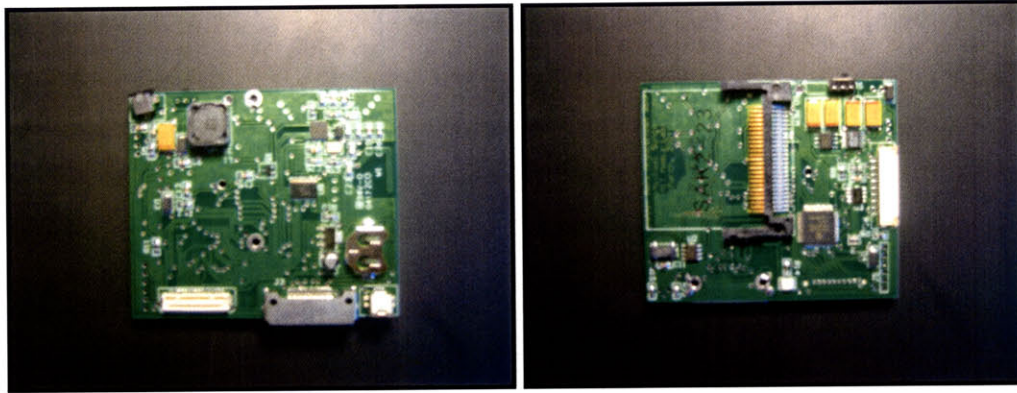
## **3.2 SAK2 Data Acquisition and Sensor Hub Board**

In order to effectively observe contextual data, a platform must have a means to gather, process, and interpret this real-time contextual data. To facilitate this, the LiveNet system includes a modular sensor hub that can be used to instrument the mobile device for contextual data gathering.

The SAK2 (Swiss-Army-Knife 2) Board is a very flexible data acquisition board that serves as the central sensor hub for the LiveNet system architecture. It was designed to be backward compatible to a previous generation data acquisition board called the Hoarder board [Gerasimov 2003]. The SAK2 board was designed primarily to interface a variety of sensing technologies with mobile device-based wearable platforms to enable real-time context-aware, streaming data applications. The SAK2 is an extremely flexible data acquisition hub, allowing for a wide variety



of custom as well as third-party sensors to interface to it. In addition to being a sensor hub, the SAK2 can also operate in stand-alone mode (i.e., without a Zaurus or mobile PDA host) for a variety of long-term data acquisition (using the CF card connector) and real-time interactive applications.



**Figure 4: SAK2 Data Acquisition and Sensor Hub Board (Front/Back).** The SAK2 is the primary interface for the LiveNet platform between the embedded mobile device and the customizable plug-and-play sensor network and supports a variety of wireless and wired communication protocols. The SAK2 has an on-board power regulator capable of providing power to the LiveNet system components as well as the sensor network during sustained monitoring applications. The SAK2 can also be used for stand-alone data acquisition applications.

## SAK2 Specifications

- PIC18F452 40-Mhz PIC microcontroller (up to 10 MIPS) 16-bit processor with 32KB FLASH program memory, 1.536 KB RAM, 256 Bytes EEPROM. This processor is equipped with a full set of analog (10-bit) and digital inputs, on-board hardware multiplier, timers, compare/capture/PWM modules, in-system flash programming, and is supported by a wide range of development tools.
- LTC 1625 Linear Technologies 5V synchronous high-efficiency step-down switching power regulator (5-36V input range). This high-efficiency power regulator accepts a wide variety of input voltage sources appropriate for wearable applications, including four AAA battery packs and Li-Ion/Li-Polymer batteries. Provides up to 2A of current at around 95% efficiency, which is necessary to support a sensor network along with the Zaurus (which can draw upwards of 1.5A assuming a depleted battery charging + WiFi

activity + full-backlight + 100% processing power utilization). The regulator can also accept unregulated power through the sensor network if no other power source is available. A 3.3V regulated power is also generated (used by the Nordic transceiver) and is also broken out to the B2B connector and sensor port.

- Nordic nRF2401 low-power 2.4 GHz wireless RF transceiver with on-board quarter-wavelength monopole antenna. Capable of megabit data rates and frequency hopping over 125 independently addressed channels over an effective range of 100 meters. This is used to access the variety of wireless sensors currently being developed at the MIT Media Laboratory.
- Compact Flash connector that accepts both Type I and Type II CF memory cards as well as IBM Microdrives.
- DS1302Z Dallas Semiconductor real-time clock with 12-mm coin battery backup (provides date and time to second resolution, for time stamping sensor data. Finer-resolution time-stamping can be generated from internal microcontroller timing.
- Sensor port connector: providing serial/I2C/regulated and unregulated power/USB signaling to MIThril-compatible sensors and peripherals (accelerometers, IR tag readers). Arbitrary numbers of sensors can be connected to the sensor network.
- Separate I2C serial port for peripherals on the I2C bus using high-insertion cycle Hirose connector. Supports up to 400 Kbits/s in (Fast-mode).
- FPC-based RS-232 connector (for interfacing with Zaurus) with an additional port connection that can be wired to a standard DB9 for PC-compatible (5V inverted) signaling. Supports up to 115.2 Kbits/s data rates.
- B2B connector for sensor/expansion daughter cards. This low-profile connector breaks out 5V and 3.3V regulated power, I2C, RS-232, and analog/digital I/O channels from the microcontroller.

- Power connector with high insertion-cycle, locking Hirose connector with high current rating. This power connector includes the 1wire interface connected to the microcontroller, to monitor battery life and usage statistics.
- Microcontroller programming port for easy in-system programming. The SAK2 can optionally be programmed on-the-fly through its serial port as well.

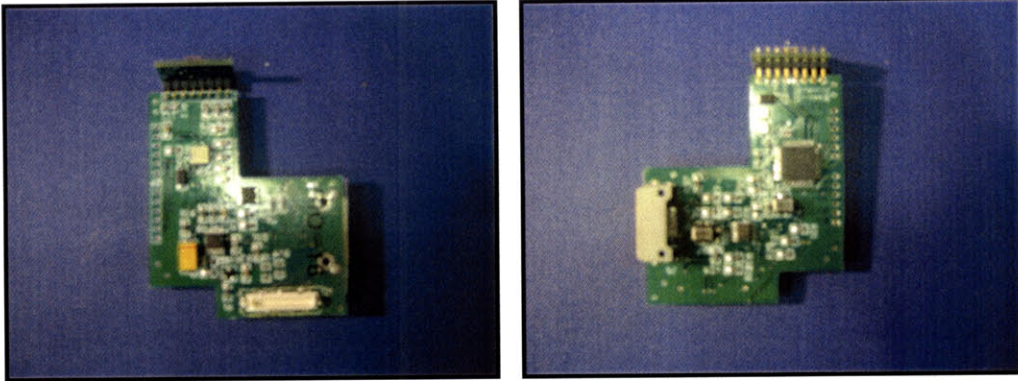
## **SAK2 Firmware**

The SAK2 firmware on the PIC452 interfaces with the host-side PDA embedded system. It provides all the buffering functionality for the dynamically configured sensor network to enable the asynchronous communications between the sensor network and mobile device. The firmware individually allocates a buffer for each configured sensor stream, and optimizes the sensor read scheduling depending on the sampling rate of each sensor.

The SAK2 firmware was written to be highly robust during operation in a wide variety of real-world ambulatory environments. In such environments, oftentimes sensors can become temporarily disconnected or power to the system can be lost if wires get jostled about from normal movement. As such, the system must be resilient to these normal rigors of wearable use. The SAK2 firmware can continue to operate correctly after being power-cycled as well as gracefully recovering from disconnections with the host device by dynamically determining settings and previous state after being reattached.

## **3.3 Physiologic and Contextual Sensing Technology**

The BioSense is a very extensible, fully-integrated physiologic sensing board that interfaces with the SAK2 via the low-profile, board-to-board connector, and is intended to support long-term health monitoring and activities of daily living applications.



**Figure 5: BioSense Daughterboard (Front/Back).** The BioSense board is designed to work directly with the SAK2 sensor hub, providing 3D motion, skin conductance, temperature, ECG/EMG sensing as well prototyping amplification circuitry for analog signals. The BioSense can also act as a bridge converter to attach other serial-based 3<sup>rd</sup> party sensors to the sensor network.

## BioSense Specifications

- PIC18F877A 20-Mhz PIC microcontroller 14-bit processor with 8KB FLASH program memory, 3 Bytes RAM, 256 Bytes EEPROM. This processor is equipped with a full set of analog (10-bit) and digital inputs, on-board hardware multiplier, timers, compare/capture/PWM modules, in-system flash programming, and is supported by a wide range of development tools.
- 3D accelerometer based on the ADXL202JE sensor capable of measuring accelerations  $\pm 2$  G in 3 axes of motion, capable of accurately identifying accelerations over most of the range of normal human motion.
- 1 ECG channel: Electrocardiogram via the INA321 differential instrumentation amplifier, capable of amplifying the potential generated by the heart (on the order of millivolts). The ECG can also serve as a single channel EMG (electromyogram) sensor for measuring muscle activity.
- 1 GSR channel: Galvanic Skin Response capable of measuring skin conductance changes over large excitation levels with tunable trimpots for tonic level adjustments (typically between 0.5 to 3 micro-siemens for a relaxed individual). This GSR was benchmarked



against gold-standard GSR equipment used by psychotherapist researchers, with a high degree of concordance.

- 3 OPA4336 operational amplifier channels: for temperature or other analog sensors (default lead set includes the LM35 precision temperature sensor capable of 0.25 °C resolution over a temperature range of -55 °C to 150 °C. Amplifiers can also be daisy-chained via jumpers to provide additional amplification stages for small-signal sensors (a piezoelectric strain respiration sensor capable of measuring sub-millimeter resolution displacements utilizes this amplifier chain).
- 4 additional general purpose analog channels that can be interfaced to a variety of other analog signal devices.
- Serial-to-I2C Converter: A preponderance of medical diagnostic devices output their data using the RS-232 protocol, which can be attached to the serial interface of a host computer. As such, the physiological sensing board breaks out the I2C and RS-232 lines, which allow us to connect any type of third-party RS-232 based sensors to the MIThril I2C sensor network with only a patch in the firmware. Currently the finger pulse-oximetry and portable blood pressure sensors have been interfaced through this pathway into our sensor network.
- High-insertion cycle, locking Hirose connector as an integrated lead set connector for ECG/GSR/sensors with virtual ground, ground, and power.
- B2B connector to attach to the SAK2 board. This low-profile connector breaks out 5V and 3.3V regulated power, I2C, RS-232, and analog/digital I/O channels from the microcontroller.
- Programming port for easy in-system programming. The SAK2 can optionally be programmed on-the-fly through its serial port as well.
- 2.5V bias reference voltage: (used by the ECG amplifier)

With the combined physiological sensing board and third-party sensors, a fully outfitted LiveNet system can simultaneously monitor and record 3D accelerometer, audio, ECG/EMG, galvanic

skin response, temperature, respiration, blood oxygen, blood pressure, heat flux, heart rate, IR beacon, and up to 128 independently channeled environmental activity sensors.

## **Other Sensors/Peripherals**

Along with the core physiological sensing capabilities of the LiveNet system with BioSense daughtercard, a whole host of other custom and 3<sup>rd</sup> party sensors can be seamlessly integrated with the system, including:

**WMSAD Board:** Wearable Multiple Sensor Acquisition Board, providing a 3D accelerometer, IR tag, IR tag readers (vertical, for in-door location in place of GPS, and horizontal for peer or object identification), and microphone for telephony-grade 8-khz audio. This is interfaced to the SAK2 via the I2C port. [Elledge 2003].

**Squirt IR Tags:** IR beacons that can broadcast unique identifiers (up to 4 independent signals from separately mounted and direction-adjustable IR-LEDS). These can be used to tag individuals, objects, locations (such as used in arrays on the ceiling to identify location to within meter resolution within indoor settings where GPS is not effective), or even as environmental sensors to identify the actuation of certain events such as opening/closing of drawers, cabinets, or doors.

**IR Tag Reader:** to be used in conjunction with the Squirt tags to identify tagged objects, people, or even locations [MIThril Sensors].

**Accelerometer Board:** 3D accelerometer board very useful for a variety of activity classification. It has been demonstrated that a single accelerometer board can be used to accurately classify activity context (standing, walking, running, lying down, biking, walking up stairs, etc). Interfaced to the SAK2 using the sensor port.

**BodyMedia SenseWear:** An integrated health sensor package which provides heart rate (via a Polar heart strap), galvanic skin response, 2D accelerometer, temperature (ambient and skin), and heat flux in a small form-factor package worn on the back of the arm. The SAK2 can interface to the SenseWear wirelessly via a 900-Mhz transceiver attached to the serial-toI2C bridge (the transceiver interface and heart rate monitor was discontinued in the SenseWear Pro 2)

**MITes Environmental Sensor:** a wireless 3D accelerometer using the nRF 2.4 GHz protocol has been developed by the house\_n group at the Media Lab for wireless environmental sensors for monitoring human activities in natural settings [Tapia et al. 2004]

**Uber-Badges:** Multifunctional boards with on-board DSP processor capable of processing audio features, RF transceiver, IR transceiver, 5x9 brightness-controllable LED output display, vibratory feedback, navigator switch, flash memory, audio input/microphone and optional LCD display. This badge is meant for social-networking experiments and other interactive distributed applications [Uber-Badge]

**Location Beacons:** A location beacon based on the Bluetooth protocol. The Class 1 bluetooth beacon's gain can effectively be set by varying the input voltage to the chipset; it was tailored to provide roughly room-area coverage (around a radius of 30 feet).

**Bluetooth Sensors:** Bluetooth sensors, including skin conductance and ECG-based sensor devices have been developed.

The LiveNet system can also be outfitted with BlueTooth, SD, or Compact-Flash based sensors and peripherals, and other I/O and communication devices including GSM/GPRS/CDMA/1xRTT modems, GPS units, image and video cameras, memory storage, and even full-VGA head-mounted displays [Icuiti].

### 3.4 Human Factors Research

Human factors considerations are an oft-overlooked area in mobile application design. To successfully implement technological tools in an educational setting, it is important that the tools' user interfaces do not require a significant amount of cognitive load that can detract from the learning experience. These tools must also be easily accessed and in the proper form factor for people to easily transport, store, and use.

One of the most critical areas when designing educational applications is the user interface, since it must be intuitive for people to effectively interact with the application to increase their productivity and/or learning ability. In addition, interfaces must be optimized for other modes of interaction appropriate for mobile devices. For example, an application requiring extensive

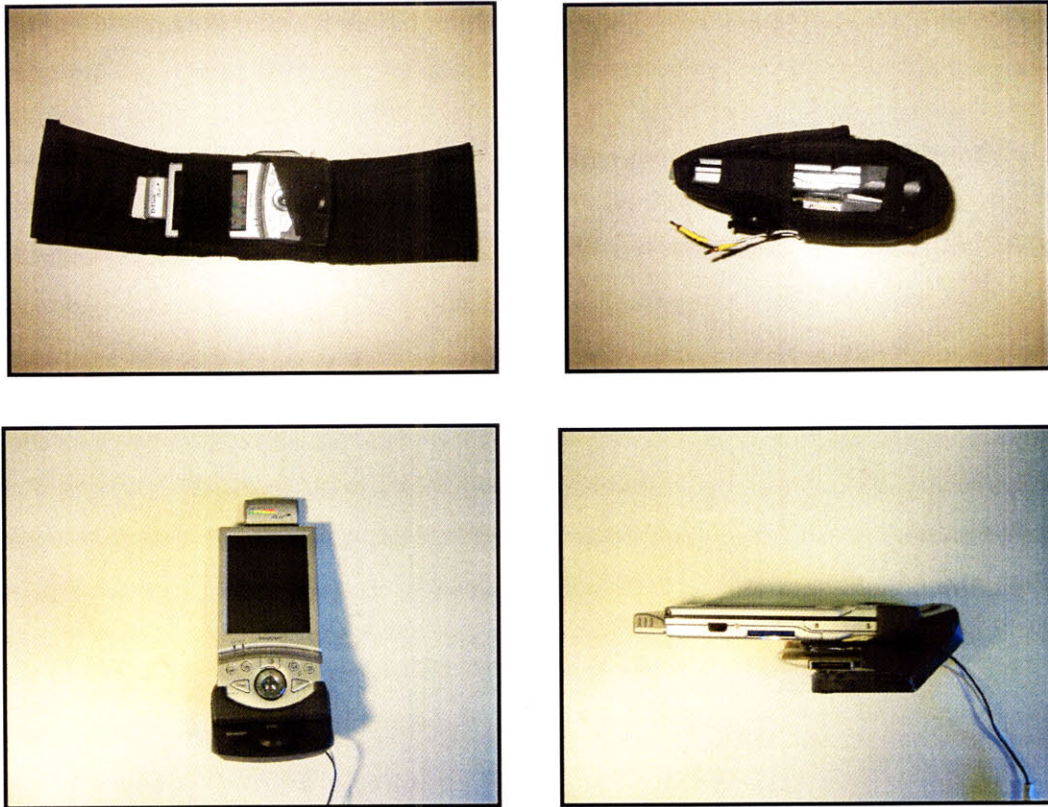
graffiti-style writing or typing using micro-keyboards while walking might limit the application's effectiveness.

As such, we have tried to adopt more novel interfaces that follow the constraints imposed by a wearable mobile system. Such constraints include limited screen real estate (a 320x240 qVGA head-mounted display or standard PDA screen), lack of pointing devices or use of mobile systems in situations where using a pointer while engaging in other tasks is impractical, limited attention and cognitive load issues (in contrast to traditional desktop interface models which demand total attention), and quick access to stored and real-time contextual information.

We have also utilized graphical display environments such as Qtopia on the Sharp Zaurus to create completely keyboard-free, button-based input devices that can be used to gather contextual data on the user's current state. These interfaces can be prompted by contextual data (such as a person's location, physiological/psychological state, movement, etc.) to interact more appropriately with the user. These context-driven interface design concepts have been implemented in many of our applications (see Section 3.6 for LiveNet application development).

The form factor and design of the hardware systems was also made to be compact and amenable to integration into a number of potential wearable systems depending on application. The SAK2 board was designed to fit in a small space behind the injection-molded plastic cradle that originally came with the Sharp Zaurus, providing a natural connection via a FPC serial connector provided by the manufacturer. This provides a completely seamless, integrated package for the SAK2 sensor hub to connect with the Zaurus. It also provides a way to undock the Zaurus from the rest of the system, potentially allowing interaction applications while having the SAK2 still buffering data from the sensor network. This data can then be resynchronized and streamed again over the WiFi through the Zaurus when it is re-docked with the rest of the system.

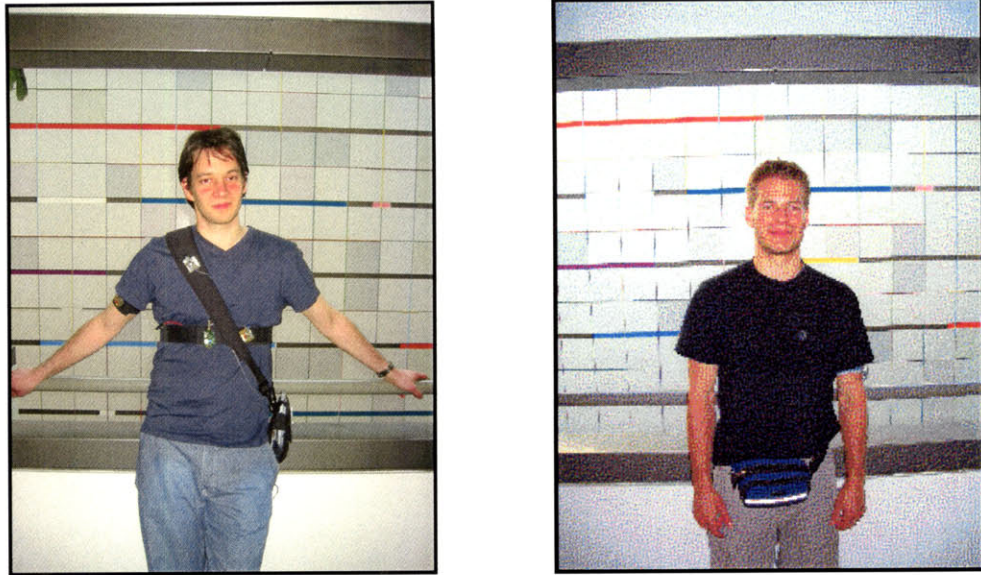
Research into a wide variety of materials and basic harness designs were tested for appropriateness for long-term monitoring use, including a variety of modified commercially-available PDA/cell-phone holsters, bandolier-style holsters, shoulder-worn bags, waist-belts, etc.



**Figure 6: Open and side profiles of a LiveNet holster package (top) and cradle assembly for easy docking of the Zaurus to the SAK2/BioSense subsystem (bottom). This system allows the Zaurus to be removed from the holster for interaction applications while simultaneously allowing the SAK2 sensor hub to buffer and store sensor data, which can be synced once the device is re-docked with the cradle.**

The final design integrates the specific needs of a flexible wearable platform, including places to put the batteries, wire straps, and places to put sensors while still being comfortable. There is a front flap release that allows the PDA device to be easily removed for interaction applications involving the screen and keyboard. The final design revolves around a shoulder strap to a holster, made of very durable ballistic nylon material and Velcro for sensor attachments. The main shoulder strap has built-in wire straps to allow wires to be run from the LiveNet holster to sensors that may make sense in the upper chest area, such as a microphone, IR transceivers, etc. This wearable rig was the primary setup used for the ARIEM Cold-Exposure Study as well as for demos such as the BorgViewer systems.





**Figure 7: Wearable options including a bandolier-style system (left) and a ‘fanny-pack’ waist strap system (right). The bandolier system was designed as a compact form factor that allows the PDA device to be quickly accessed for interaction applications. Sensors such as the audio microphone and IR beacons can be conveniently attached to the strap on the chest. The waist strap system was more comfortable for women and more convenient for long-term wear such as the MGH Depression Study.**

An alternative waist strap system was also developed for more long-term use. The waist strap system was more comfortable for women and more convenient for long-term wear since the pouch can be repositioned depending on whether the patient was sitting or lying down. This system was the primary one used in the MGH Depression Study.

### 3.5 Software System Infrastructure

Even though the variety of applications available for modern mobile devices is quite compelling, they are typically standalone programs with little flexibility or extensibility. In typical out-of-the-box systems, there are critical problems in distributed inter-process communication, signal processing, and sensor data classification that were not addressed by operating systems or currently available software tools.

To address these problems, the MIT Wearables Laboratory has developed a software architecture [DeVaul et al. 2003] to combine the best features and practices from a range of research systems and methodologies, doing so in an open, modular, and flexible way. The three important software

systems that form the foundation of this architecture are: The Enchantment Whiteboard system for inter-process communication, the Enchantment Signal system for high-bandwidth data streaming, and the MIThril Real-Time Context Engine machine learning infrastructure. This software infrastructure ties everything together, allowing network-transparent streaming data communication to arbitrary endpoints capable of real-time, context-aware interaction. These tools address critical needs in the development of mobile applications while imposing minimal constraints on the nature of these applications.

Enchantment is an implementation of a whiteboard inter-process communications and streaming data system suitable for distributed, light-weight embedded applications. It provides a uniform structure and systematic organization for the exchange of information that does not require synchronous communications. Enchantment is intended to act as a streaming database, capturing the current state of a system (or person, or group) and can support many simultaneous clients distributed across a network and hundreds of updates a second on modest embedded hardware. It has even been demonstrated that one can use the Enchantment for bandwidth-intensive VoIP-style audio communications. The flexible system architecture allows the LiveNet platform to be configured as either a completely distributed system as well as a centralized client-server system, depending on the design tradeoffs between resource consumption and fault tolerance.

The Enchantment Signal system in addition to the Enchantment Whiteboard System, provides an intuitive infrastructure for producing and consuming heterogeneous streams of information. The Signal system, in conjunction with the Enchantment Whiteboard, allows us to simply plug in a variety of sensors, and have those streams of information simply flow from producer to consumer in a network-transparent way. In addition, very high data bandwidths are achievable, and can easily scale up to the bandwidth limits of the underlying wireless 802.11 network.

The MIT Wearables Laboratory has also developed a high-level framework called the MIThril Context Engine [DeVaul & Pentland 2003] for applying statistical machine learning techniques to the modeling and classification of body-worn sensor data. The important design features of the system are simplicity, modularity, flexibility, and implementability under tight resource constraints.

The Context Engine abstracts the data analysis into distinct steps, including the transformation of raw sensor data into features more suitable for the particular modeling task, the implementation of statistical and hierarchical, time-dependent models that can be used to classify a feature signal in real time, and the development of Bayesian inference systems which can use the model outputs for sophisticated interpretation and decision-making.

More details of each software subsystem is discussed in turn below.

## **The Enchantment Whiteboard**

The Enchantment Whiteboard system is an implementation of a whiteboard inter-process communications system suitable for distributed, light-weight embedded applications. Unlike traditional inter-process communications systems (such as RMI, Unix/BSD sockets, etc.) which are based on point-to-point communications, the Enchantment Whiteboard is based on a client/server model in which clients post and read structured information on a whiteboard server. This allows any client to exchange information with any other client without the attendant  $\Theta(n^2)$  complexity in negotiating direct client-to-client communication; in fact, this is possible without knowing anything at all about the other clients. This does for inter-process communication what web browsers and web servers do for document publishing – it provides a uniform structure and systematic organization for the exchange of information that does not require synchronous communications.

The Enchantment Whiteboard goes beyond the web server analogy by allowing clients to subscribe to portions of the whiteboard, automatically receiving updates when changes occur. It allows clients to lock a portion of the whiteboard so that only the locking client can post updates. And it even supports symbolic links across servers, allowing whiteboards to transparently refer to other whiteboards across a network. The Enchantment Whiteboard is also lightweight and fast, imposing little overhead on the communications.

The Enchantment Whiteboard is intended to act as a streaming database, capturing the current state of some system (or person, or group) and on modest embedded hardware can support many simultaneous clients distributed across a network and hundreds of updates a second. We have even demonstrated the ability to use the Enchantment Whiteboard with the Signal system for bandwidth-intensive VoIP-style audio communications.



## **Enchantment Whiteboard server**

Every Enchantment Whiteboard application requires that at least one Whiteboard server be running on an accessible and known host. The server is generally run as a daemon on the wearable device as well as on other more general-use computers. Enchantment applications on the wearable that need to communicate amongst themselves usually use the local server, and applications that do external resource discovery can contact remote servers.

Each Enchantment Whiteboard server instantiation maintains a global tree that clients can access by using the Enchantment Whiteboard client API. While the server has the ability to save its tree state to persistent media between sessions, Whiteboard data is generally time dependent and perishable. We have found the current, single-threaded version of the server to run stably on various types of Unix systems, including Linux and Solaris.

Producers of information (e.g. motion data from an accelerometer-enabled LiveNet system) can thus publish handles to data, and consumers (e.g. a server running a motion classification application) can subscribe to that data by looking on the whiteboard for the appropriate handles. Applications can access the Enchantment Whiteboard server through the Enchantment Whiteboard client API. The library routines require a Universal Node Locator (UNL), which is analogous to a web URL, to address a particular node on a specific Whiteboard server. A UNL that specifies the node “motion” that is below the node “classifiers” on the “localhost” server would read “localhost:/classifiers/motion”. When a UNL points to a node that has children, library routines generally operate on the node and all its descendants.

## **Enchantment Signal System**

For higher bandwidth signals, especially those related to the sharing and processing of sensor data for context aware applications, we developed the Enchantment Signal system. The Signal system is intended to facilitate the efficient distribution and processing of digital signals in a network-transparent manner. The Signal system is based on point-to-point communications between clients, with signal “handles” being posted on Whiteboards to facilitate discovery and connection. In the spirit of Whiteboard interactions, the Signal API abstracts away any need for signal producers to know who, how many, or even if, there are any connected signal consumers.

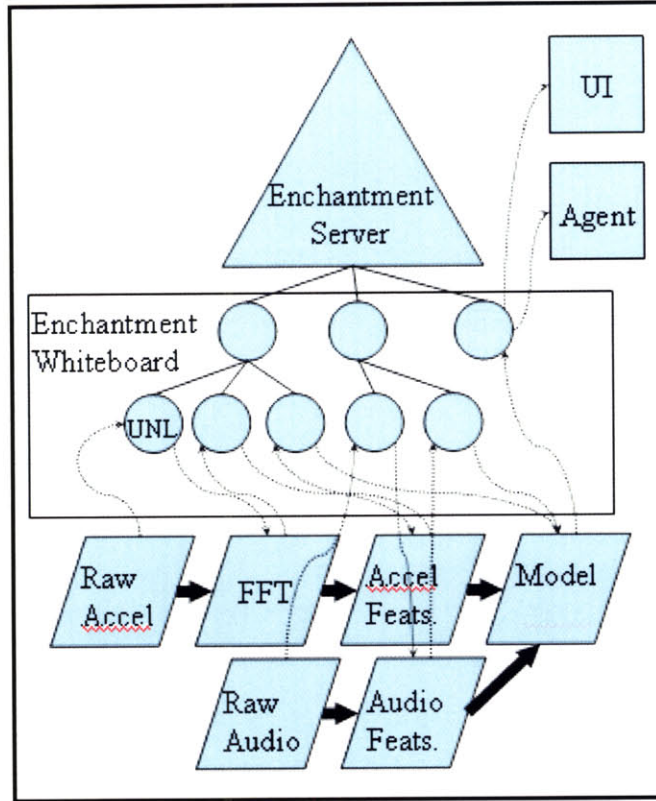


Figure 8: Typical LiveNet application showing interaction between the Enchantment Server, Enchantment Whiteboard, Enchantment Signal System, and MIThril Real-Time Context Engine. Producers of data (such as an accelerometer or microphone device) can publish UNL handles to a particular type of data on the Enchantment Whiteboard, which is moderated by the Enchantment Server. The Enchantment Signal System the forwards these data streams, called signals, (depicted by arrows) to feature extraction modules, which then subsequently post their own output as UNLs that other modules downstream can subscribe to. A model can then use the features extracted for classification, and the context results can then be sent to the front-end agent and UI modules.

Any type of structured numeric data can be encoded as a signal. Signal producers may be sensors, feature extractors, filters, or regression systems, and may produce other signals in turn. A typical organization is a sensor signal producer talking to a feature extraction signal consumer, which in turn produces a feature signal that is consumed by one or more modeling or regression systems.

The results of modeling or regression can themselves be signals, or (more typically) posted on a whiteboard for other clients to use. This model allows for the implementation of principled statistical machine learning systems, as described in the next section.

## MITHril Real-Time Context Engine

The MITHril Real-Time Context Engine is a practical, modular framework for the development of real-time context classifiers and models. After using the LiveNet sensing infrastructure to sample and collect data from the real world, we follow a standard machine learning process flow to pre-process, analyze, and model the data so that we can create useful inferences. The MITHril Real-Time Context Engine abstracts the process of feature extraction, modeling, and inference stages, as shown in Figure 9 below:

### **Machine Perception Stages:**

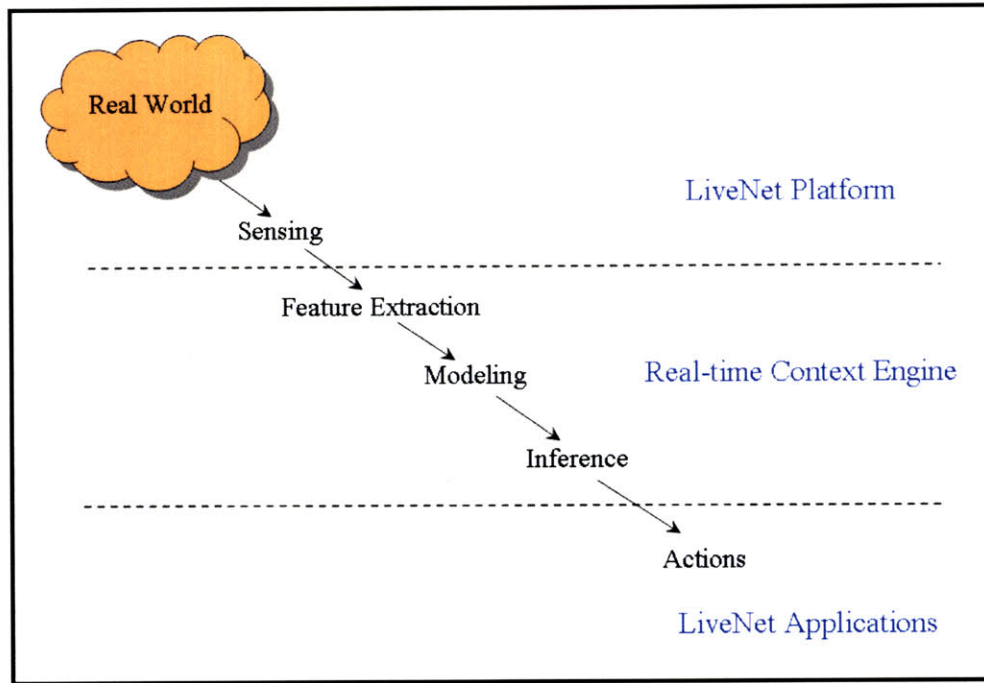
**Feature Extraction:** In the feature extraction stage, a raw sensor signal is transformed into a feature signal more suitable for a particular modeling task. For example, the feature extraction stage for a speaker identification classification task might involve converting a sound signal into a power spectrum feature signal.

**Modeling:** In the modeling stage, a statistical model (such as a Gaussian mixture model or hidden Markov model) is used to classify a feature signal in real time. For example, a Gaussian mixture model could be used to classify accelerometer spectral features as walking, running, sitting, etc.

**Inference:** In the inference stage, the results of the modeling stage, possibly combined with other information, are fed into a Bayesian inference system for complex interpretation and decision-making.

When tackling a new problem, much of the feature extraction, modeling, and inference groundwork is first prototyped offline using flexible analysis tools such as Matlab and Kevin Murphy's Bayes Net Tool Kit [Murphy 2001], a powerful graphical models and Bayesian inference system. We have been able to implement a number of the more commonly used modules in C and C++ using the Enchantment Signal and Whiteboard system, such as real-time spectral feature generators, and feature generators specifically tailored for voice features described in Section 4.3. The process outlined above assumes that one already knows what features to extract and the type, structure, and parameters of the model. In order to learn these parameters, one must first gather a certain amount of labeled training data, do appropriate

analysis to choose features, and then learn the model parameters. The Real-Time Context Classifier toolkit provides data logging, labeling, and analysis tools to facilitate this process.



**Figure 9: Process flow for developing context-aware applications. Real-world events can be detected using the sensing infrastructure of the LiveNet system. This data must first be pre-processed to extract useful features amenable for modeling. These features can then be used as inputs to train the models that we select to apply to the situation of interest. These models can then be used to provide inferences by LiveNet applications to develop context-aware applications to classify the current context of the system or even predict the future state of the system.**

### 3.6 LiveNet Application Development

The LiveNet system was designed with extreme flexibility in mind. With the software architecture API, it is possible to rapidly configure a system to and even prototype a variety of applications that involve distributed sensing, real-time information processing and data streaming. The following examples show the breadth of the type of applications that can be developed using the LiveNet platform.

A few of these applications were developed specifically to facilitate data collection for the studies involved with this thesis. Other applications were proof-of-concept systems made to demonstrate



the ability of the platform to support a variety of potentially interesting applications. A number of these applications are discussed in turn.

## BioRecord: GUI-Based Multimodal Sensor Acquisition and Annotation

BioRecord is a front-end GUI interface to a multimodal sensor data acquisition and annotation system that allows a researcher to simply click the appropriate checkboxes to configurable sensor selection and automated data acquisition (chosen among sensors such as multiple accelerometers, audio, skin conductance, ECG, temperature, and polar heart rate). Data is stored in a separate folder on the SD memory card indicated by the test session and individual files are stored consecutively by a concatenation of the starting date and timestamp.

The application also allows one to annotate and timestamp events and notes to the physiology, which is important later in having precise timing for analysis purposes and reducing the typical headaches associated with manually synchronizing the sensor data to each other or to annotations.

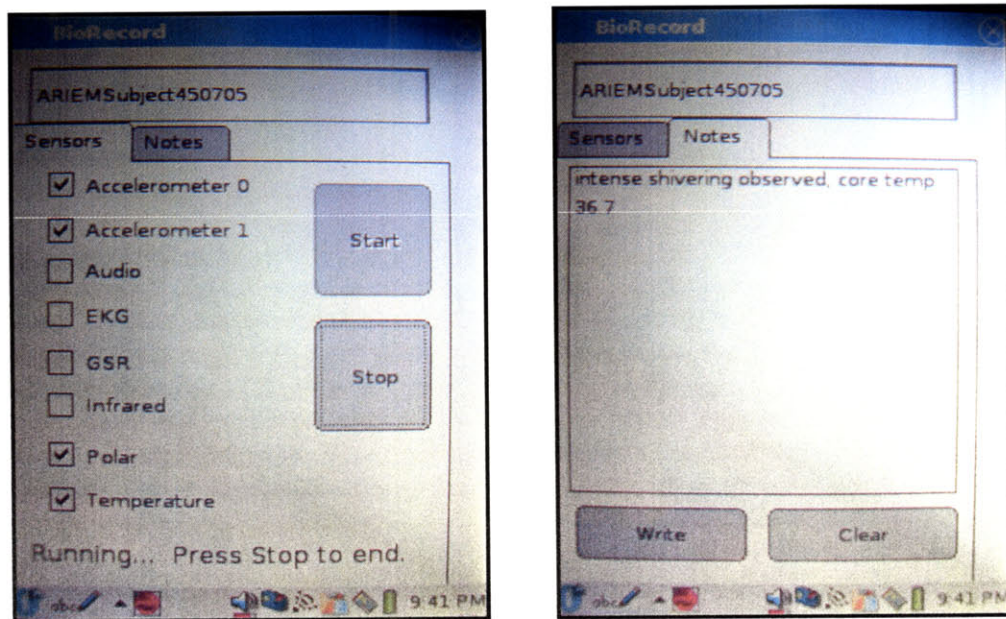


Figure 10: BioRecord application showing the sensor selection screen (left) and the timestamp and annotation screen (right). This GUI-based application is a front end to the Enchantment infrastructure, allowing researchers to use turn-key LiveNet systems to easily record and annotate a wide variety of sensing data for experiments.

BioRecord was used at the ARIEM site, as well as a number of the other collaborations discussed in Section 3.7 below to allow researchers to easily collect and annotate data.

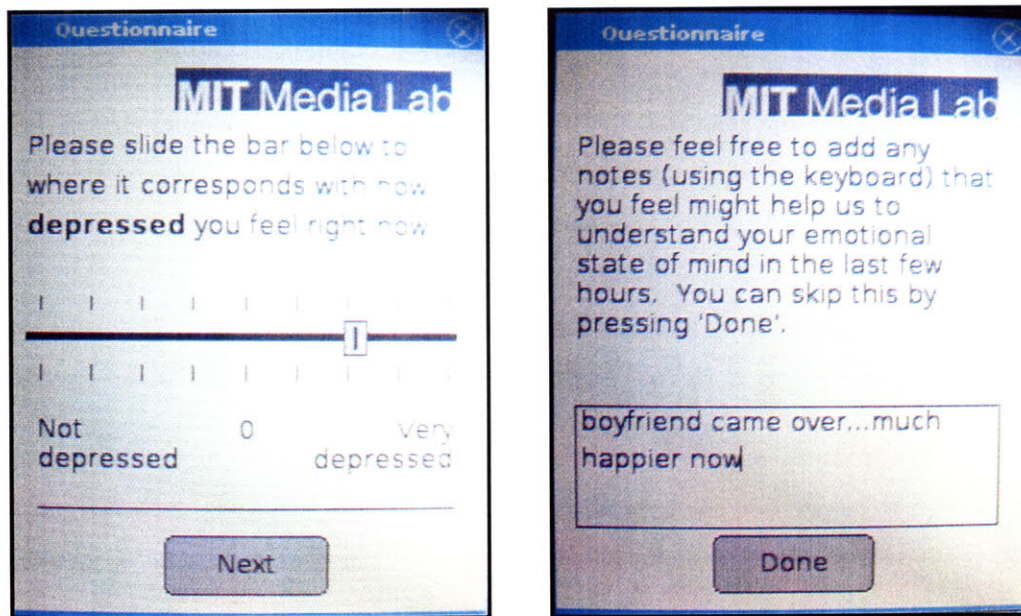
### **Quest: Configurable Experience Sampling Application**

In addition to monitoring and recording sensor data, it is oftentimes very important to be able to assess the immediate contextual state by querying the user. A popular assessment tool for studying the activities of people in natural settings is the experience sampling method (ESM), whereby a subject carries an alarm-enabled device that “samples” the user’s contextual state by initiating a questionnaire on some predetermined schedule [Larson & Csikszentmihalyi 1983]. These types of ESM surveys can potentially be less susceptible to subject recall errors than other self-report recall surveys.

Towards this end, a configurable questionnaire application called Quest was developed. Quest is made to be extremely flexible to configure and easy to use, so as to be applicable in a variety of remote ambulatory domains where the patient is not in contact with study personnel. The questions, dialog, and comments sections can all be simply configured using a set of text files that are placed in a local directory, allowing for the Quest to be reconfigured with ease. In addition, the program can be configured to auto-start any number of times during the day as well as sound various programmable alarms in order to remind the user of various actions, such as to take a survey or to initiate an activity. A couple of screen shots of the application configured for a survey for the MGH Depression study are show below:

The Quest application can even be configured to auto-start based on the contextual state or contextual changes of the user as sensed by the sensor network attached to the LiveNet system. By having a context-aware experience sampling capability, the system can ask more relevant questions based on the actual contextual state of the user, as opposed to just activating based on a predetermined schedule [Intille et al. 2003]. This allows the system to acquire more information about a situation of interest by just sampling during or just after the context changes or is detected. For example, the Quest program can be configured to activate when a user suddenly moves around, a typical indication when there is a change in contexts. Thus, the system can be programmed to only activate a questionnaire during a context change if that is what the

experimenter is interested in, thus avoiding constant interruption annoyance of the user in a more standard experience sampling methodology necessary to capture a certain level of detail.



**Figure 11: Quest application showing the visual analog scale survey question (left) and subject comments sections (right). This customizable survey application can be programmed to auto-start based on timed intervals or even event-specific triggers from the sensor network or other inputs.**

## **PatientMonitor: Robust Automated Monitoring System**

For long-term ambulatory monitoring in naturalistic settings, it is oftentimes inconvenient to have to manually configure a data acquisition program before the start of a monitoring session. This complicates the data collection protocol and leaves room for human error in miss-configuring the settings. In fact, such was the case with the MGH Depression Study described in Chapter 7, with study staff that were not well versed in using PDAs, as well as the fact that they did not have the time nor inclination to wait for the PDA to boot up, start and configure the data collection settings before even starting the protocol of strapping up the patient. In these instances, even having a GUI-based application such as BioRecord for data acquisition is not appropriate.

In order to reduce the interaction necessary to initiate daily data collection, the PatientMonitor application was created. This reduced the protocol to simply turning on the PDA for it to configure the appropriate sensors (hard-coded) upon boot-up of the LiveNet system. Thus, a

monitoring session with an arbitrarily sophisticated start-up script can be programmed and initiated with the touch of a button. This allows a subject to put on and use the system themselves without the need for outside support and oversight.

Although the software infrastructure has been extensively tested, occasionally problems can occur that stop the system from further broadcasting sensor data. In addition, the PDA embedded operating system can be notoriously unreliable, causing reboots, freezes, etc. For long-term ambulatory monitoring applications, the ability for the system to automatically check periodically for proper operation is critical. PatientMonitor anticipates these types of problems that can occur during long-term monitoring applications where there is nobody to reset the system. The system continually checks things such as whether the internal software infrastructure server is up or whether the sensors are still giving valid data, etc. every few minutes and restarts processes that are not functioning properly. This allows the system to be robust and continue to record valid data despite a variety of problems that can occur during ambulatory monitoring situations, including power disconnects, PDA operating system glitches, and sensor meta-stabilities.

### **PatientTracker: Location Beacon Person Tracker**

In recent years, the Bluetooth protocol has gained momentum and many modern mobile devices device have begun to incorporate the technology to wirelessly connect to each other. Other researcher have shown the ability to use this fact to track and monitor people's social interaction and location [Eagle 2005].

A similar application called PatientTracker has been developed specifically for the MGH Depression Study. This application allows a LiveNet system, enabled with a standard CF-based Bluetooth card, to scan the environment for bluetooth devices in its immediate vicinity and record the Bluetooth IDs that it comes across. This application is set to periodically scan (the scan interval was adjustable, but typically an interval of a few minutes was optimized to be able to track the high level behavior of a person's location while extending the battery life of the PDA). Since these Bluetooth IDs are unique, they can be used to refer to static locations (such as areas that have fixed devices or Bluetooth beacons installed), or they can be mobile devices that people carry that can be effectively used to determine their identity. Thus, the PatientTracker allows the



LiveNet system to track the location of a person over time, and identify social patterns by identifying people who come into contact with the user.

We were able to modify a Class-1 Bluetooth device that was designed by the Responsive Environments Group at the MIT Media Lab to develop a location beacon to identify if a patient moves within a particular space. These beacons are relatively low-cost and have a very small form factor that can be hidden behind objects in a room, and can be deployed in large quantities to cover a large physical space such as a hospital ward. The location beacons are completely adjustable in range, with an effective radius between ten to over several hundred feet, so that a beacon could be tailored to various sized spaces. These location beacons can be used to track the high level movement patterns of an individual, much like companies such as Living Independently [Living Independently].



**Figure 12: Bluetooth Location Beacon and associated housing and power supply. These beacons allow the LiveNet system to track the location of a patient within living spaces. The location beacons are completely adjustable in range, with an effective radius between ten and over several hundred ft.**

PatientTracker can be configured in two ways. The first is by placing a LiveNet system scanner (in a simple PDA form-factor) in the locations of interest. This enables the tracking system to be completely transparent to a user, assuming that the person usually carries around a mobile phone

anyway. Alternatively, the Bluetooth beacons can be scattered around the living area and other locations of interest if the person is already being monitored and wearing a LiveNet system.

## **Real-Time Physiometric Displays**

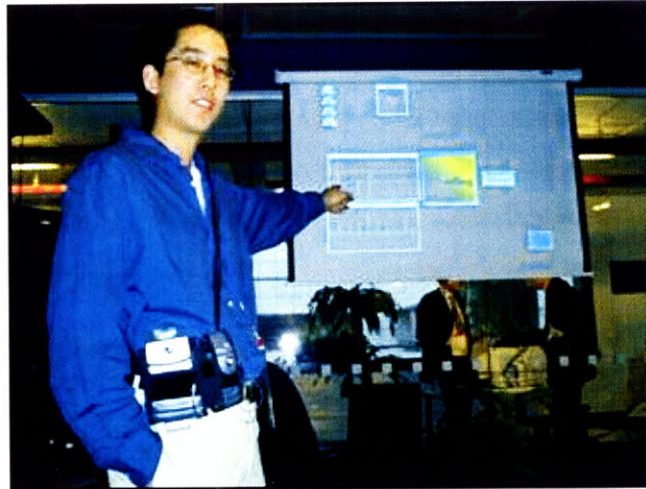
We have used the LiveNet platform to create proof-of-concept interfaces for integrating physiology information into real-time biofeedback and visualization displays. Consider a soldier monitoring command center, whereby each soldier's multimodal physiology can be streamed in real-time to a display. A display can also aggregate all the soldiers' physiological signals in real-time, allowing people to get a sense of the overall status of a large group of people. The soldiers' aggregated psychophysiology statistics can be used to immediately gauge the overall high-level status of a platoon. Each individual soldier also has the option of viewing his own physiological signals on their own LiveNet systems, which is useful as biofeedback for the speaker to identify his own mental or physical state.

We have deploying this type of application in experimental sessions within live seminar classes being taught at the Media Lab, where it is useful to monitor individual and group reactions to structured interactions. For the classes, the speaker and audience members were outfitted with LiveNet systems that can measure skin temperature, EKG, and galvanic skin response. The systems are also configured to act as an accelerometer-based real-time head nodding/shaking classification systems [Sung & Gips 2003] One can imagine how the aggregated physiology (which may show interest levels, for example) can be used to time-stamp specific points in a lecture which may be of interest, and audio snippets from those periods of time can be compiled as a digest of the most interesting points of a talk. Figure 13 shows the author wearing a LiveNet system streaming real-time data to a nearby projector.

The data can be used to gauge interest and agreement levels in real-time, and to cross-compare self-reporting results to baseline information such as unconscious nodding in agreement and psycho-physiologic cues such as heart rate and galvanic skin response, which are highly correlated with stress and sympathetic nervous system arousal.

The real-time visualization display can serve a dual role for the speakers as well as teacher. The audience's aggregated psychophysiology statistics can be used to gauge audience attention and interest in the form of socio-biofeedback. The speaker can also be wired with his own

physiological signals, which is useful as biofeedback for the speaker to identify his own mental state, or by the teacher, who can observe the effects of the speaking or negotiation interaction. Within the context of a structured interaction of a speaking or negotiation class, the person's performance can be studied in a very controlled environment, giving the individuals as well as the instructor valuable feedback.



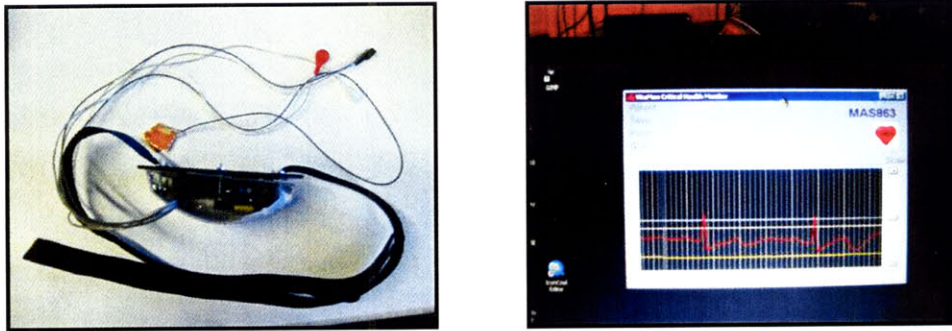
**Figure 13: LiveNet system performing real-time FFT analysis and classification on accelerometer data, and wirelessly streaming real-time EKG/GSR/temperature and classification results to a remote projector. We can easily imagine a system such as this being used to stream distributed data to a variety of endpoints for real-time soldier or critical medical monitoring systems.**

### **VitaMon: Low-Cost Critical Health Monitors**

In developing countries, it is difficult to provide adequate monitoring services to the critically ill, as hospitals are understaffed and cannot afford expensive monitoring systems. This application focuses on using the LiveNet system to develop a portable, wearable monitoring device with low-power processing and sensing capabilities that can monitor and record vital patient signs (specifically EKG, temperature, and galvanic skin response) in order to take the place of direct human monitoring. In the event that a patient's vital signs indicate an emergency (examples include the heart rate dropping below a certain rate, temperature above/below accepted norms, and hyper-ventilation and/or shock as indicated by the skin conductance sensor detecting a ramp in conductance over the baseline in a specified time interval), the monitoring device can sound an alarm to alert the attention of medical staff. The device can also record the patient's vital signs



onto local storage on the wearable monitor, which can be reviewed by a doctor at a later point [Sung 2002].



**Figure 14: VitaMon monitor and sensors (left) and real-time critical physiology from being streamed to a remote laptop for storage and visualization** The VitaMon is a very low-cost (sub \$30 in volumes) physiological event monitor capable of real-time data streaming based on the SAK2 data acquisition board.

The VitaMon wearable monitor is a proof-of-concept device that can allow ambulatory remote healthcare monitoring. The entire system is simply a combination of the SAK2 data acquisition board configured in stand-alone operation with the BioSense board. These systems can be manufactured in large volume for \$10-20, very amenable to low-cost systems for third-world countries. By incorporating a PDA or other mobile device, we can visualize the real-time data streams.

## **BorgViewer: Distributed Real-Time Data Communication System**

The LiveNet platform is very capable of this distributed group applications involving streaming real-time information and context results. LiveNet systems can also be configured to enable point-to-point real-time data streams, functioning as communicators that can provide instantaneous contextual information to a distributed group of people. Through these instantaneous communication channels, it is possible to create the feeling of virtual proximity of peer groups, despite the fact that there may be large geographical separation between individuals.

In order to demonstrate the potential of using LiveNet for real-time data streaming over a distributed network of mobile devices, we have developed an application called BorgViewer. The application integrates an interface that can be used to instantaneously visualize the real-time data

and context streamed from remotely located people over a distributed network. In this system, individual people, or “borgs”, each wear a LiveNet system, which monitors ECG, movement, skin conductance, audio, and any other sensing which may be attached to their sensor network.



**Figure 15: A “borg” wearing a LiveNet system using the BorgViewer application to stream raw accelerometer and ECG signals as well as to display the real-time context classification information. A user running BorgViewer can instantaneously visualize the real-time data of a borg and also get a sense of its immediate context based on the real-time activity classifier and proximity information of other nearby borgs based on IR beacons. This distributed network of LiveNet systems is perfect for command-and-control and other group-based applications.**

Each person in the system can use their GUI interface to pick another borg and instantly view their real-time physiology and context, which can be computed locally and visualized on their mobile device. In addition, each borg has a tagged IR beacon that serves as a personal identifier. By using the IR receiver (using the WSMAD board, for example), one can view a particular borg’s location and interaction patterns with other borgs or objects that are tagged with IR beacons.

Thus we demonstrate the potential for real-time data streaming, as even high-bandwidth data such as audio can be used to provide voice over IP (VoIP) style communications. We can do this in infrastructure mode over an existing 802.11 network, or even configure the systems to operate in

ad-hoc mode over their own private network. Each individual borg is uniquely tagged with its own IP address, and information can be encrypted via the SSL capabilities of Enchantment infrastructure to ensure private communications for sensitive information.

It is not hard to see how we can generalize the BorgViewer application to command and control applications, such as a platoon of soldiers who need to communicate with each other and immediately identify an individual's status to coordinate activities to work in a cohesive manner.

### **3.7 LiveNet Collaborations**

The LiveNet systems have been used in a variety of domains, including all the studies involved with this thesis as well as a number of other outside collaborations:

#### **Physiology and Violence in Games Study, Harvard Medical School**

This study is conducted by the Harvard Medical School to study the effects of violence in video games on the psychology of video game players. The researchers have developed a script that can identify biases toward violence based on the reaction times to certain words. Using scripts, they can measure the violence biases before violent gaming exposure and afterwards. LiveNet systems were used to record the multimodal physiology and behavior of gamers while they played a variety of violent video games. By collecting the physiology data, one can study how the violent video games affect physiology and behavior, and potentially correlate to violence biases. This study demonstrates how third-party researchers can easily collect physiology data using turn-key LiveNet systems with applications such as BioRecord.

#### **ECG Shirt/Non-invasive Sensing Technologies, Franhauser Institute**

The Franhauser Institute for Reliability and Microintegration (IZM) is in collaboration to develop micro-miniaturized ECG sensing hardware based on the BioSense board. The collaboration is focused on integration technologies for wearable electronics and electronics in textiles. Specifically, technologies for encapsulation, flexible substrates, and textile processing are used to create completely integrated wearable sensing systems. This miniaturized sensing board is intended to be integrated into an ECG shirt with fabric-based conductive electrodes, thereby reducing the burden of the more invasive gel-based electrodes that need to be attached to the body

with adhesives. Created prototypes have shown that it is possible obtain a relatively noise-free ECG signal based on the BioSense design.

### **Pervasive Healthcare Applications, National Taiwan University**

The Ubicomp Lab at National Taiwan University has also developed a collaboration to use the LiveNet platform in conjunction with their smart home infrastructure. The LiveNet systems are used as ambulatory wearable monitoring devices, in order to obtain physiology information that would otherwise not be available based solely on sensing technologies embedded into the environment. The goal is to develop light-weight sensing systems for health monitoring that can communicate with infrastructure embedded in the environment for telemedicine applications. This collaboration demonstrates the potential of integrating the LiveNet system in conjunction with pervasive technologies to create powerful context-aware systems for pervasive healthcare applications.

### **Eldercare Monitoring, British Telecom**

Eldercare will become an increasingly important topic over the next few decades as demographics around the world shift towards older populations. In order to reduce the strain that these aging populations will cause on the healthcare infrastructure, it is necessary to develop proactive healthcare technologies that will enable the elderly to live independently. This can be done by providing long-term monitoring systems that can identify the long-term behavioral patterns of individuals as they go about their daily lives.

British Telecom is using LiveNet technologies to develop long-term ambulatory health monitors that can potentially identify activities of daily living (which have been shown to correlate with high-level functioning and even mortality of the elderly). In addition, by looking at long-term behavioral patterns based on location tracking, it is hoped that systems can be developed to identify deviations from normal behavior which can be flagged as warnings, to which healthcare providers can detect and respond.

## **Parkinson's Disease Study, Universidad Politécnica de Madrid**

We have also started collaborating with the Universidad Politécnica de Madrid on a study to develop an automated Parkinson's symptom detection system. Continuous monitoring of physiologic and behavioral parameters may be extremely effective in tailoring medications to the individual Parkinson's patient. In Parkinson's patients, there are a variety of symptoms and motor complications that can occur, ranging from tremors (rhythmic involuntary motions), akinesia (absence or difficulty in producing motion), hypokinesia (decreased motor activity), bradykinesia (slow down of normal movement), and dyskinesia (abnormal or disruptive movements).

For these patients to function at their best, medications must be optimally adjusted to the diurnal variation of these symptoms. In order to optimize treatment, the managing clinician must have an accurate picture of how a patient's symptoms fluctuate throughout a typical day's activities and cycles. In these situations, a patient's subjective self-reports are not typically very accurate, so objective clinical assessments are necessary. By using the LiveNet system's wearable accelerometers, the researchers hope to collect movement data and use machine learning algorithms to classify the movement states of Parkinson's patients and provide a timeline of how the severity of the symptoms and motor complications fluctuate throughout the day.

## **Epilepsy Study, University of Rochester Center for Future Health**

A collaboration has also been initiated with the University of Rochester's Strong Hospital to characterize and identify epileptic seizures through accelerometry and to develop an ambulatory monitor with a real-time classifier using the LiveNet system. Typically, epilepsy studies focus on EEG and EMG-based physiology monitoring. However, as demonstrated by activity classification studies, accelerometry is a powerful context sensor that can be applied to the domain of epilepsy. The study protocol is currently being designed, and we hope to start subject runs in Spring, 2006.

A challenge of the study is the fact that the epileptic seizures of a particular individual can manifest themselves in an extremely wide range of idiosyncratic motions, in contrast to Parkinson's' patients, whose movements typically follow distinct, characteristic motions. However, motions from the epileptic seizures of a particular individual are normally fairly



consistent. As such, a motion classification system specifically tailored to a particular individual could be highly effective at being able to identify an epileptic seizure. Many times, the epileptic individual has no recollection of a seizure, so a system that could determine if a seizure has occurred could be very useful for doctors to be able to properly diagnose the patient or to develop applications to alert caregivers.

## **Soldier Health Monitoring, U.S Army ARIEM Natick Laboratories**

Army Rangers and other soldiers must perform physically and mentally demanding tasks under challenging environmental conditions ranging from extreme heat to extreme cold. Thermoregulation, or the maintenance of core body temperature within a functional range, is critical to sustained performance. A joint research collaboration with the Army Research Institute for Environmental Medicine (ARIEM) at the Army Natick Labs was initiated to study the effects of harsh environments on soldier physiology through the use of non-invasive sensing. Specifically, non-invasive accelerometer sensing was used to determine hypothermia and cold exposure state, as part of a broader initiative to develop a physiologic monitoring device for soldiers under the US. Army's Objective Force Warrior Program.

A real-time wearable monitor was developed using the LiveNet system that is capable of accurately classifying shivering motion through simple accelerometer sensing and statistical machine learning techniques [Duda & Hart 2000]. Based on the collected data, we demonstrate that shivering can in fact be accurately distinguished (95% accuracy) from general body movements in various activities using only continuous accelerometer sensing. Results also indicate that specific modes of shivering may correlate with core body temperature regimes, as a person is exposed to cold over time. In fact, preliminary results over six subjects show that we can triage a soldier into three core body temperature regimes (Baseline/Cold/Very Cold) with accuracies in the 92-98% range using HMM (Hidden Markov Models) modeling techniques. This exploratory research shows promise of eventually being able to develop robust real-time health monitoring systems capable of classifying cold exposure of soldiers in harsh cold environments with non-invasive sensing and minimal embedded computational resources. This study is discussed in detail in Chapter 6.

## Depression Study, Massachusetts General Hospital

Mental diseases rank among the top health problems worldwide in terms of cost to society. Major depression, for instance, is the leading cause of disability worldwide and in the U.S. Depressive disorders affect approximately 19 million American adults and have been identified by both the World Health Organization and the World Bank as the second leading cause of disability in the United States and worldwide [Kessler et al. 2003, Murray & Lopez 1996].

Towards understanding the long-term biology associated with severe depression, we have recently initiated a pilot study to assess the physiological and behavioral responses to treatment of major depression in subjects in an inpatient psychiatric unit prior to, during, and following electroconvulsive therapy (ECT). This study, the first of its kind, intends to correlate basic physiology and behavioral changes with depression and mood state through a 24-hour, long-term, continuous monitoring of clinically depressed patients undergoing ECT. We are using non-invasive mobile physiologic sensing technology in combination with sensing devices on the unit to develop physiological and behavioral measures to classify emotional states and track the effects of treatment over time. This project is a joint collaboration with the Massachusetts General Hospital (MGH) Department of Psychiatry.

The goal of this study is to test the LiveNet system based on the known models of depression and prior clinical research in a setting with combined physiologic and behavioral measures with continuous ambulatory monitoring. Changes in these measures have been shown to correlate with improvements in standard clinical rating scales and subjective assessment following treatment for depression throughout the course of hospitalization. In the future, these correlates may be used as predictors of those patients most likely to respond to ECT, for early indicators of clinical response, or for relapse prevention. This study is discussed in detail in Chapter 7.

## Chapter 4      **Non-Invasive Features and Modeling**

The LiveNet platform has the flexibility to accommodate a wide variety of sensors, but the research in this thesis focuses only on using minimally-invasive physiology sensing that can also act as general context sensors for capturing behavior. This Chapter is dedicated to the description of these non-invasive sensors and the features that have the most promise to be used in context inference. Specifically, these sensors are accelerometers, audio, temperature/heat flux, heart rate, and skin conductance.

These easily obtained sensing parameters can be correlated to more involved physiological measurements such as ECG, blood pressure, pulse oximetry, respiration, four of the most important diagnostic measures for medicine. All of these physiological sensors available to the LiveNet system can allow us to better quantify behavior and verify our models of affective/physiological state. However, it is demonstrated that some of the more invasive measures are highly correlated with features that can be derived from the smallest, easily obtainable non-invasive sensors.

The goal is to be able to characterize and classify even context states, perhaps even soft variables such as interest, attention, using only measurements of the aforementioned non-invasive sensing. We first discuss some general features that are applicable across all types of sensor data. We then discuss the significance and specifics of each type of non-invasive sensing in turn, beginning with the least invasive features. This is followed by a discussion of the methodologies for combining and reducing the available feature sets to find the most useful features. We then conclude with a section on the various modeling techniques that we employ to develop the classifiers and models that we use in the studies in Chapter 5-7.

### **4.1      Generic Physiology Features**

A variety of potential features for the physiological information have been explored for descriptive and classification power. Some of these are standard statistical features, some are pulled from signaling theory generally applicable to the domain of physiological sensing, and others are combinations thereof. A variety of other features are motivated by the underlying

physiological principle (such as using frequency domain analysis for semi-periodic signals). Each of these features is discussed in turn below.

The most general features are simple statistical measures, such as the mean and standard deviation of the data over a specified window. Consider a raw signal, where  $n = 1, \dots, N$  are the samples in the window of interest and  $X_n$  is the value of the  $n^{\text{th}}$  sample. The mean of the signal gives a sense of the center of the distribution of the signal, or the central tendency of the variable, and can be calculated as follows:

$$\mu_x = \frac{1}{N} \sum_{n=1}^N X_n \quad 4-1$$

The standard deviation of the signal gives a sense of the spread, or variation in the signal:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (X_n - \mu_x)^2} \quad 4-2$$

It is also very important to be able to normalize a raw data signal. A typical normalization technique is to center the data by subtracting the mean and scaling by the standard deviation, as follows:

$$\tilde{X}_n = \frac{X_n - \mu_x}{\sigma_x} \quad 4-3$$

$\tilde{X}_n$  is the normalized signal with zero mean and unit variance, also known as the standardized or z-scored version of the original raw signal. This is particularly useful to be able to compare different features, as the underlying raw physiology features can have arbitrary scaling relative to each other.

$\delta_x$  is the mean of the absolute values of the first differences of the raw input signal, which approximates an aggregate gradient:

$$\delta_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |X_{n+1} - X_n| \quad 4-4$$

$\tilde{\delta}_x$  is the mean of the absolute values of the first differences of the normalized signal.

$$\tilde{\delta}_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1} - \tilde{X}_n| = \frac{\delta_x}{\sigma_x} \quad 4-5$$

$\gamma_x$  is the mean of the absolute values of the second differences of the raw signal, which allow use to get quantify the acceleration or deceleration in a signal:

$$\gamma_x = \frac{1}{N-2} \sum_{n=1}^{N-2} |X_{n+2} - X_n| \quad 4-6$$

$\tilde{\gamma}_x$  is the mean of the absolute values of the second differences of the normalized signal

$$\tilde{\gamma}_x = \frac{1}{N-2} \sum_{n=1}^{N-2} |\tilde{X}_{n+2} - \tilde{X}_n| = \frac{\gamma_x}{\sigma_x} \quad 4-7$$

The features represented by Equations 4-1 through 4-7 represent standard statistical features that have been referenced and used in emotion physiology literature [Vyzas & Picard 1998, Vyzas & Picard 1999, Picard et al. 2001]. Note that these types of statistical measures are easily computable using a sliding window, and are particularly amenable for real-time calculations [Vyzas & Picard 1999].

We also include a number of standard features from signaling theory, as the derived physiology features are in fact signals. The energy of a signal is a measure of the magnitude of the signal. This can be calculated as the “area under the curve” of the squared signal as follows:

$$E(x) = \sum_{i=1}^N x_i^2 \quad 4-8$$

Another useful measure is the entropy of a signal. In information theory, the entropy is equivalent to the amount of randomness in a signal, defined by Claude Shannon as follows:

$$H(x) = -\sum_{i=1}^N p(i) \log_2 p(i) \quad 4-9$$

The autocorrelation of a signal is the cross-correlation of the signal with itself. It is useful for finding repeating patterns in a signal or identifying the fundamental frequency of a signal, which may not actually contain that frequency component, but implies it with harmonic frequencies.

The discrete autocorrelation  $R$  at lag  $j$  is given as:

$$R(j) = \sum_{n=1}^N (X_n - \mu_x)(X_{n-j} - \mu_x) \quad 4-10$$

All of the features discussed so far are time domain measures. In addition to these features, many of the physiology signals contain frequency information that can be extracted using signal processing techniques such as Fourier analysis. This type of analysis is based on the Fourier transform, an integral transform that re-expresses a function in terms of sinusoidal basis functions. This is a very common signal processing technique to decompose a signal into its component frequencies and their amplitudes [Oppenheim 1989]. The formula for a continuous Fourier transform is:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad 4-11$$

The Fourier transform can compute the frequency content of the original input signal and output a weighted spectrum. However, it is not practical to use the continuous Fourier transform for analysis. Thus, we utilize the discrete version of the Fourier transform, called the discrete Fourier transform (DFT). The DFT can be expressed as follows:

$$F(n) = \sum_{k=0}^{N-1} f_k e^{-i\omega kn/N} \quad k = 0, \dots, N-1 \quad 4-12$$

While computing a DFT directly requires  $O(n^2)$  operations, the fast Fourier transform (FFT) algorithm can compute the discrete Fourier transform in  $O(n \log n)$  operations, making it very practical for computer systems to be able calculate these transformations. We have demonstrated that embedded hardware, such as the Sharp Zaurus used by LiveNet, has no problem computing an FFT in real-time using optimized FFT code [Frigo & Johnson 2005].

Using Fourier analysis, we can easily calculate the power spectral density of a signal and use these spectral magnitudes as features. If the Fourier transform is calculated over successive windows over the input signal, we can generate a spectrogram which depicts the frequency content of a signal over time. We use these spectrograms extensively in order to be able to intuitively visualize and discriminate between various types of movement, as discussed in the accelerometry section below. Other features of interest can be constructed by binning the frequency response to find averages over particular frequency bands. Specifically for heart rate data, heart rate variability measures using frequency domain analysis is very common, as discussed in Section 4.5.

Typically before we perform a Fourier analysis on the data, we apply a Hanning window to preprocess the raw data. This is necessary to avoid aliasing artifacts that occur if the start and end points of the data do not match. In the Hanning window, the center value is 1 (no change to the data), while the end values are zero. A Hanning window at position  $k$  and length  $N$  is defined by the following function:

$$h(k) = \frac{1}{2}(1 - \cos(2\pi k / N)) \quad 4-13$$

We can also use a Hanning window as a simple filter to smooth out a raw signal, as is done for the raw skin conductance and heart rate described below.

While the above features are quite universally relevant, they do not take into account the particular underlying physiology, and as such, do not exploit information which may be applicable in developing useful features. There is also no normalization for long-term variation of these signals (for example, base tonic levels for skin conductance can vary day to day). In the

following sections, we discuss the features related to each type of specific physiology sensing data in turn.

## 4.2 Accelerometry/Motion

One of the most important pieces of contextual information is the activity context of an individual. Being able to predict an individual's immediate activity state is one of the most useful sources of contextual information. For example, knowing whether a person is driving, sleeping, or exercising could be immensely useful for a health wearable to calculate general energy expenditure or to initiate a proactive action. Many studies on motion-based activity classification have been conducted because of its relative importance to context-aware systems.

Specifically in the realm of gerontology, being able to determine activities of daily living is key to assessing the status of the elderly, as it has been shown to be a good predictor of nursing home acceptances and mortality [Wiener et al. 1990]. As such, there is a large research effort to identify new technologies that are able to automatically track and classify activities of daily living [Lynch et al. 2003]. With low-burden sensing requirements, such as the activity context classification study described in Chapter 5, it becomes practical to capture a person's long-term activity and health context, which can then be used to detect potentially important deviations from normal healthy behavior.

There are a variety of features that can be derived from motion information that is very useful for a variety of classification tasks. Raw acceleration data can be used to identify posture, given that there is always a gravitational acceleration component that can be isolated assuming the sensors are clearly oriented. Thus, simple accelerometry can easily tease out information such as lying down or standing. Mean acceleration features can identify activities such as going up and down the escalator or traveling in vehicles.

Spectral features in particular can oftentimes be used to identify between activities. Spectral entropy and other variance measures are highly correlated with activity intensity. In addition, many activities are repetitive in nature, and the frequency of this repetition can be extracted using spectral analysis. Examples include walking, shivering, riding a bicycle, and vibrations from traveling in certain types of vehicles. We have found that spectral features of accelerometer data are particularly useful in context classification. We have demonstrated use of these features for



activity classification in a variety of domains, including soldier shivering classification [Sung et al. 2004] and Parkinson's disease classification [Klapper 2003].

In the past, simple accelerometer-based activity context (standing/walking/running) and head-nodding/shaking agreement classifiers were developed [Sung & Gips 2003]. Preliminary evidence suggests that discriminating between these activity contexts based solely on general body accelerometry spectral features is possible. In this thesis, we generalize and extend this work to include the classification of a rich variety of activity contexts beyond basic movement such as walking/running to include a variety of context states such as lying down/sleeping, going up and down stairs and elevators, riding in vehicles, working at the office, watching T.V., falling down, exercising, etc. This research is important as it indicates that it is feasible to do activity classification on embedded hardware without any specialized setup, wires, or other unwieldy parts. Furthermore, the sensors do not have to be placed in any particular orientation to work since we use spectral features of the overall acceleration magnitude on the torso of an individual. This general activity context classification study is discussed in Chapter 5.

The MIT Wearable Computing Group has a custom-designed accelerometer device that is capable of measuring acceleration in all three spatial axes, using a part by Analog Devices (the ADXL202) that is commonly used by research groups.

### **4.3 Social Interaction and Speech Features**

While speech recognition is the standard goal of most voice processing applications, audio sensors have also demonstrated power to be used as environmental sensors capable of identifying immediate contextual state [Siuru 1997, Saint-Arnaud 1995]. An important piece of context, for example, is the ability to determine if a person is talking to other people or socializing, or perhaps actuating events such as a door closing, etc.

In many situations, non-linguistic social signals (body language, facial expression, tone of voice) are as important as linguistic content in predicting behavioral outcome [Ambady & Rosenthal 1992, Nass & Brave 2004]. In a wide range of situations (marriage counseling, student performance assessment, jury decisions, etc.) an expert observer can reliably quantify these social signals and with only a few minutes of observation predict about a third of the variance in behavioral outcome (which corresponds to a 70% binary decision accuracy) [Ambady &

Rosenthal 1992]. In fact, Ambady et al. show that even ‘thin-slices’ of these social signals of just a few seconds can be just as accurate in predicting outcomes as long-term windows of analysis. This is particularly amazing given that the observation of social signals within such a ‘thin slice’ of behavior can predict important behavioral outcomes (divorce, student grade, criminal conviction, etc.) when the predicted outcome is sometimes months or years in the future.

Social interaction has commonly been addressed within two different frameworks. One framework comes from cognitive psychology, and focuses on emotion. Ekman and Friesen are the most well-known advocates of this approach, which is based roughly on the theory that people perceive others’ emotions through stereotyped displays of facial expression, tone of voice, etc. [Ekman & Friesen 1977]. This simplicity and perceptual grounding of this theory has recently given rise to considerable interest in the computational literature [Picard 1997]. However, serious questions about this framework remain, including the question of what counts as affect? Does it include cognitive constructs such as interest or curiosity, or just the base dimensions of positive-negative, active-passive? Another difficulty is the complex connection between affect and behavior: adults are skilled at hiding emotions, and seemingly identical behaviors may have different emotional roots.

The second framework for understanding social interaction comes from linguistics, and treats social interaction from the viewpoint of dialog understanding. Kendon et al. [Kendon et al. 1975] and Argyle [Argyle 1987] are among the best known pioneers in this area, and the potential to greatly increase the realism of humanoid computer agents has generated considerable interest from the computer graphics and human-computer interaction community [Cassell & Bickmore 2003]. In this framework prosody and gesture are treated as annotations of the basic linguistic information, used (for instance) to guide attention and signal irony. At the level of dialog structure, there are linguistic strategies to indicate trust, credibility, etc., such as small talk and choice of vocabulary. While this framework has proven useful for conscious language production, it has been difficult to apply it to dialog interpretation, perception, and for unconscious behaviors generally.

We can employ a new computational framework, that of social signaling, in which speaker attitude or intention is conveyed through the amplitude and frequency of prosodic and gestural activities [Pentland 2003]. This framework is based on the literature of personality and social

psychology, and is different from the linguistic framework in that it consists of non-linguistic, largely unconscious, signals about the social situation, and different from the affect framework in that it communicates social relation and not speaker emotion. It is different in another way as well: it happens over longer time frames than typical linguistic phenomena or emotional displays, treating gestures more like a motion texture than individual actions, and it appears to form a largely independent channel of communication. In the language of the affect framework, these signals are sometimes identified by the oxymoronic label 'cognitive affect', whereas in the linguistic framework they might be related to dialog goals or intentions. Social signaling is what you perceive when observing a conversation in an unfamiliar language, and yet find that you can still 'see' someone taking charge of a conversation, or establishing a friendly interaction [Gladwell 2000, Gladwell 2005].

Among these the types of social signaling, tone of voice and prosodic style two of the most discriminating, even though (and perhaps because) these signals are sent without people being consciously aware of them [Nass & Brave 2004]. Pentland constructed a number of measures based on vocal social signaling, including activity level and stress [Pentland 2003, Pentland-B 2004, Pentland et al. 2004]. We have identified four general types of vocal social signaling, namely activity level, engagement, stress, and mirroring which have strong support in voice analysis and social science literature.

Activity level can be determined by looking at measures such as speaking vs. non-speaking percentage of speaking time. A robust voicing segmentation algorithm developed by Basu was used to be able to identify voicing segments [Basu 2002]. Voice activity can be determined through a two-level HMM to segment an audio speech stream of a person into voiced and non-voiced segments. These voiced segments can then be combined into longer speaking and non-speaking segments [Basu 2002]. Voice activity can then be calculated as the percentage of speaking time (relative to non-speaking time) and the frequency of the voiced segments. These feature extraction algorithms have been demonstrated to be robust even in settings with background noise. Conversational level is calculated as the z-scored percentage of speaking time plus the frequency of voiced segments.

Stress in the voice can be measured by the variation in speaking characteristics of an individual, either purposefully for prosodic emphasis or unintentionally as a manifestation of psychological

stress through physiological changes. For every voiced segment, the mean energy, spectral entropy, and fundamental formant is extracted, and longer time windows can be used to calculate the standard deviations of these measures. These measures all reflect the degree of variation in the voice, which has been correlated with speaker stress. The z-scored sum of these standard deviations is taken as a measure of speaker stress; such stress can be either purposeful (e.g., prosodic emphasis) or unintentional (e.g., physiological stress caused by discomfort). These features have been shown to be correlated to a variety of phenomenon, including negotiation ability and interest [Pentland et al. 2004, Madan et al. 2004].

Engagement can be measured by the influence each person has on the other's turn taking. When two people are interacting, their individual turn-taking dynamics influence each other and can be modeled as a Markov process [Jaffe et al. 2001]. By quantifying the influence each participant has on the other, we obtain a measure of their engagement and whether they were driving the conversation. To measure these influences, we model their individual turn-taking by an Hidden Markov Model (HMM) and measure the coupling of these two dynamic systems to estimate the influence each has on the others' turn-taking dynamics [Choudhury 2003, Choudhury & Pentland 2003].

Mirroring behavior, in which the prosody of one participant is 'mirrored' by the other (oftentimes unconsciously), is considered to signal empathy, and has been shown to positively influence social interactions such as the outcome of a negotiation [Chartrand & Bargh 1999]. We can create mirroring features in which we measure frequency of short interjections at time scales less than one second (e.g., 'uh-huh'), and back-and-forth exchanges typically consisting of single words (e.g., 'OK?' 'OK!' 'done?' 'yup.').

The above-derived features are extracted from the speech of a sole individual and represent their overall speaking style as measures of speaker emphasis and prosody. However, it is also possible to extract features calculated collectively from a set of audio signals (or some form of combined audio), which represent the dynamics of interaction between individuals. This can be done by looking at the influence parameters and the number of short back-and-fourth exchanges.

To describe the interactions between people, we have developed a generalized model called the Influence Model, based on Coupled Hidden Markov Models (CHMMs) which can be formulated

with a minimalist set of parameters limited to the inner products of the individual Markov chains [Basu et al. 2001]. This allows a simple parameterization in terms of the *influence* each person has on the other. These influence parameters express then how strong the overall state is for an individual depending on the state of the other person. In this case, we use a simple two-state model of speaking vs. not-speaking to model individual dynamics, and then measure the influence parameter to determine the coupling between speakers.

It is noteworthy that all the required speech feature extraction can be implemented in real-time on existing PDA hardware. This allows us to deal with low-bandwidth speech features, rather than the raw audio signals. This has implications in data storage (audio data takes up a lot of disk storage space) as well as for privacy concerns, since speech features have the added benefit of being completely non-reversible in terms of being able to extract conversational content.

Whereas speech recognition typically requires a close-talking, high fidelity microphone, for the features of interest for contextual sensing, it is possible to use cheap microphones. Audio data for extracting voice features can be obtained by using a standard omni-directional electret microphone that is clipped to the chest of the subject being recorded.

In this thesis, we explore the possibility of using audio in this manner to determine the ambient audio texture that can be used to help determine immediate contextual state. Voicing features such as pitch, energy, rate of speaking, percentage of time speaking, and measures of participants' prosody and the dynamics of their turn-talking can be used without recording speech or extracting words or speech content to quantify conversation dynamics. The following is a list of the twenty-two voice features used in our analysis that are derived from the above background research:

**Voice Feature 1 – Average length of voiced segment (ALVS):** This is the average length, in seconds, of a voice segment in the frame of interest. This gives a sense of how long a continuous, unbroken voiced segment is.

**Voice Feature 2 – Average length of speaking segment (ALSS):** This is the average length, in seconds, of a speaking segment in the frame of interest. This gives a sense of the speaking structure in a frame (i.e. whether a person talks a little at a time or in long continuous segments).

**Voice Feature 3 – Fraction of speaking time (FST):** This is simply a ratio of the speaking time to the total time in the frame. This is a good behavioral measure of the amount of speaking per unit time that is present in the audio data.

**Voice Feature 4 – Voicing rate (VR):** This is the number of voiced regions per second speaking. Through the speaker segmentation algorithm described above, voicing regions are combined into speaking segments. This measures the average number of voicing regions per speaking segment.

**Voice Feature 5/6 – Mean/standard deviation of the fundamental formant frequency (MFF/SFF):** A formant is a peak in the acoustic frequency spectrum resulting from the resonant frequencies of the voice, measured in Hz. The formant of the lowest frequency is called the fundamental formant. We can easily pinpoint the fundamental formant frequency by performing an FFT on the audio voice data. In layman's terms, the mean of the fundamental formant indicates the pitch of the voice, and the standard deviation gives a sense of the pitch variation in the voice.

**Voice Feature 7/8 – Mean/standard deviation of the confidence in the fundamental formant frequency (MCFF/SCFF):** We can also look at the confidence interval of the fundamental formant frequency as an alternative measure to the standard deviation of the variation of pitch in the voice.

**Voice Feature 9/10 – Mean/standard deviation of the spectral entropy (MSE/SSE):** The spectral entropy of the voice is simply the entropy of the frequency components after passing the voice data through an FFT. This is another alternative measure of the randomness, or variation in the voice.

**Voice Feature 11/12 – Mean/standard deviation of the value of largest autocorrelation peak (MVACP/SVACP):** By performing an autocorrelation on the voice data, we have a time-domain method for identifying the formant frequencies of the voice. The value of the largest autocorrelation peak identifies how strong the autocorrelation peak is; i.e. how defined the fundamental frequency is, as implied by the harmonics of the voice audio data.

**Voice Feature 13/14 – Mean/standard deviation of the location of largest autocorrelation peak (MLACP/SLACP):** The location of the autocorrelation peak identifies the location of the

shift that creates the largest autocorrelation peak. As such, it is another way to identify the fundamental frequency implied by the harmonics of the voice audio data.

**Voice Feature 15/16 – Mean/standard deviation of the number of autocorrelation peaks (MNACP/SNACP):** The number of autocorrelation peaks gives an indication of the number of different harmonics that exist in the voice data, as it identifies the number of periodic signals that is buried within the data.

**Voice Feature 17/18 – Mean/standard deviation of the energy in frame (MEF/SEF):** The energy in the frame is simply the energy of the frequency components in the audio data after performing an FFT.

**Voice Feature 19/20 – Mean/standard deviation of the time derivative of energy in frame (MTEF/STEF):** We can pass the energy of the frame feature through a time derivative feature, to get the rate of change of the energy in a frame. This gives a sense of the dynamics of the voice energy through frames.

**Voice Feature 21 – Conversation Level (CL):** Conversational level is calculated as the z-scored percentage of speaking time plus the frequency of voiced segments.

$$CL = FST + \frac{1}{ALVS} \quad 4-14$$

**Voice Feature 22 – Stress Level (SL):** The sum of the standard deviations of the mean-scaled measurements for speaking energy, fundamental formant frequency, and spectral entropy, as follows:

$$SL = SEF + SFF + SSE \quad 4-15$$

## 4.4 Temperature/Heat Flux

Non-invasive skin/body temperature is another interesting measure that is explored in this thesis. Temperature and heat flux for the LiveNet system is derived from two armband sensors that effectively measure skin and ambient temperature. Using these sensors, it possible to record the

temperature profile (along with other physiology measures) of an individual for long periods of time.

Temperature is one of the most basic measures of proper body functioning, which is why a mother first checks a baby's temperature when it is behaving erratically or feeling unwell. Body temperature has had a great deal of diagnostic power in a person's general health, such as in the obvious context of fevers or thyroid disease [Schachter 2000]. It is also clear that metabolism, physical exertion, and general energy expenditure is closely linked to body temperature and heat flux.

Whereas clinical research is focused on gold-standard calorimetry measurements, it has been demonstrated that non-invasive physiology can be used to make reasonably accurate approximations [Jakicic et al. 2004]. The goal is not to produce clinically accurate metabolism measures, but to get a general sense of metabolism and physical exertion that can be used as a general contextual sensor for high-level behavioral monitoring. In particular, metabolism estimates from a combination of temperature/heat flux measurements are ideal for weight management applications or general lifestyle trending.

It turns out that heat flux is a very useful signal in identifying context switches of people's daily behavior. Typical context switches involve activity, often times in the form of walking or other exertion. These context changes can easily be detected in spikes in heat flux as a person's body warms up due to physical activity. Heat flux differentials can also occur simply from changing physical environments, such as moving from a warm building into the cold outside air.

## **4.5 Heart Rate and Heart Rate Variability**

Heart rate is probably one of the most important and fundamental physiologic measures, and has a direct correlation with much of the body's basic physiology, including heart function, overall health, physical exertion, and emotional response. Heart activity also ends up being very useful measure in behavior, as it is a reliable indicator of an individual's overall activity level; for example, heart rate accelerations occur in response to exercise, loud noises, sexual arousal, and mental effort [Fridja 1986]. It is also important to note that there is a positive correlation between heart rate and blood pressure (systolic/diastolic). Heart rate is one of the primary variables in cardiac output, which is proportional to blood pressure.



Heart rate can be measured these days through a variety of very non-invasive means. For this thesis work, the LiveNet system focuses on using a commercially-available chest strap to obtain our heart rate data, which can be accurately correlated to actual ECG information. These days, heart rate data can be very accurately and non-invasively obtained with commercially available exercise monitoring equipment such as from Polar, Inc. This type of technology gives us the ability to leverage off the diagnostic power of this type of physiological information in a variety of health and behavioral contexts.



**Figure 16: Non-invasive heart rate chest strap for detecting inter-beat interval of the heart. These types of sensors are very robust in the presence of large physical motion and have been shown to be highly correlated to more invasive ECG monitoring equipment. Most of the heart rate features of interest (such as heart rate variability measures) do not require ECG morphology information, so a non-invasive chest strap is just as ideal for heart rate sensing.**

For semi-periodic signals such as heart rate, the raw heart rate information comes naturally. The Polar heart strap naturally detects inter-beat intervals (IBIs) by recording the time between pulses, so the data produced is directly applicable to the various heart rate and heart rate variability measures discussed below. The instantaneous heart rate is simply the inverse of the IBI measurements. The raw heart rate signal  $H$  can be smoothed by convolving the signal with a Hanning window  $h$  as follows:

$$b = H * h$$

4-16

The heart rate mean can then be calculated as:

$$r_1 = \frac{1}{N} \sum_{n=1}^N b_n \quad 4-17$$

Standard heart rate features such as the mean, energy, entropy, max/min ratios can be easily extracted directly from the sensor data. The mean of the first difference of the heart rate signal can be used to estimate the average acceleration or deceleration of the heart rate:

$$r_2 = \frac{1}{N-1} \sum_{n=1}^{N-1} (b_{n+1} - b_n) = \frac{1}{N-1} (b_N - r_1) \quad 4-18$$

Heart rate variability (HRV) refers to the natural beat-to-beat variations in heart rate. Even at rest, the heart rate of healthy individuals exhibits periodic variation. This rhythmic phenomenon, known as respiratory sinus arrhythmia (RSA), fluctuates with the phase of respiration – RSA causes the heart rate to accelerate during inspiration, and decelerate during expiration. RSA is predominantly mediated by respiratory gating of parasympathetic activity to the heart; this activity occurs primarily in phase with expiration and is absent or attenuated during inspiration. Reduced HRV has thus been used as a marker of reduced parasympathetic activity.

The major reason for the interest in measuring HRV stems from its ability to predict survival after heart attack. Derived measures of heart rate variability have been shown to be correlated with a variety of conditions and health functions, most notably as a predictor for post-heart failure mortality and general health [TFESC 1996, Carpeggiani et al. 2004, Adamson et al. 2004, Kristal-Boneh et al. 1995]. In recent years, heart rate variability has enjoyed serious interest for a variety of other physiological, psychological, and psychosocial conditions ranging from depression, to anxiety, stress, and panic [Balogh et al. 1993, Kawachi et al. 1995, Yeragani et al. 1993].

It also appears that HRV varies with psychosocial factors. Several studies have now suggested a link between negative emotions (such as anxiety and hostility) and reduced HRV. Kawachi et al. reported a cross-sectional association between anxiety and reduced HRV [Kawachi et al. 1995]. Offerhaus observed lower HRV in individuals who were "highly anxious" according to the Minnesota Multiphasic Personality Inventory [Offerhaus 1980]. Yeragani et al. have published a

series of reports indicating reduced HRV (using both time domain and spectral measures) among DSM-III diagnosed panic disorder patients [Yeragani et al. 1990, Yeragani et al. 1993]. Sloan et al. reported reduced high-frequency power among 33 healthy volunteers who scored high on the Cooke-Medley Hostility scale [Sloan-A et al. 1994]. The association between negative affect and reduced HRV may thus provide a potential mechanism linking chronic stress to disease outcomes.

Time domain HRV analysis measures changes in heart rate over time or intervals between successive cardiac cycles. The calculated time domain variables can be simple, such as the mean RR interval, mean heart rate, or the difference between the longest and shortest RR interval. There are also statistical indices including IBI derived directly from the intervals themselves (such as SDNN, SDANN, and SD), or instantaneous HR and intervals derived from the differences between adjacent NN intervals (RMSSD and pNN50) [Maklik 1995, Kleiger et al. 1992]. SDNN, the standard deviation of NN intervals, is a global index of HRV and reflect all the long-term components and circadian rhythms for variability during the recording period (typically a 24-hour via a Holter monitor). SDANN is an index of the variability of the average of 5-min intervals over 24 hours. This provides long-term information, but is sensitive to low frequencies like physical activity, changes in position and circadian rhythm. SD is generally considered to reflect the day/night changes of HRV. RMSSD and pNN50 are the most common parameters based on interval differences. The pNN50 represents the fraction of NN intervals that differ by more than 50 milliseconds from the previous NN interval and RMSSD represents the root-mean-square of successive differences of NN intervals. These measures correspond to short-term HRV changes and are not dependent on day/night variations. They represent alterations in autonomic tone that are mediated by the parasympathetic response. The pNN50 measure has proven very useful in providing diagnostic and prognostic information in a wide range of conditions [Mietus 2002].

Frequency domain HRV measures require a mathematical transformation of heart rate or inter-beat intervals (IBIs) into power spectral density (PSD). These measures are commonly used as non-invasive tests of integrated neurocardiac function, since it can help distinguish sympathetic from parasympathetic regulation [Ori et al. 1992]. HRV and PSD analyses have been used to monitor a variety of pathological states [Ori et al. 1992], and predict mortality following myocardial infarction [Kleiger et al. 1978], and congestive heart failure [Saul et al. 1988] and during coronary angiography [Saini et al. 1988]; they have also been used to determine risk of

rejection after cardiac transplantation [Binder et al. 1992]. Recent work has suggested that HRV and PSD analysis also can be used to characterize a number of psychological illnesses, including major depression and panic disorders [Yeragani et al. 1991, Yearagani et al. 1993] providing a potential link between emotional states and HRV. In addition, a number of studies assessing HRV following mental stress and a recent study looking at hostility have all reported increased sympathetic and decreased parasympathetic activity [Sloan-A et al. 1994]. Decreased parasympathetic tone has been reported following acute myocardial infarction [Rothschild et al. 1988], hypertension [Malliani et al. 1991] and heart failure [Saul et al. 1988]. These findings may explain why mental and emotional stress has been identified as an independent risk factor in cardiac death following acute myocardial infarction [Frasure-Smith 1991] and may predict the risk of developing hypertension [Markovitz et al. 1993].

The HRV power spectrum contains two major components: the high frequency (0.18-0.4 Hz) component, which is synchronous with respiration and is identical to RSA. The second is a low frequency (0.04 to 0.15 Hz) component that appears to be mediated by both the vagus and cardiac sympathetic nerves. The power of spectral components is the area below the relevant frequencies. The total power of a signal, integrated over all frequencies, is equal to the variance of the entire signal. The low to high frequency spectra ratio feature is often interpreted as a measure of the relative sympathetic to parasympathetic activity of the autonomic nervous system. The low frequency power is generated mainly from sympathetic activity, with baroreceptor modulation being a major component in low frequency power. The high frequency power, in contrast, is derived from vagal, or parasympathetic activity, which is modulated by respiration. Thus, low to high frequency ratio represents a good indicator of sympathetic-vagal balance, and is used to assess the balance of the autonomic nervous system. A recent study by Sloan et al. [Sloan et al. 1994] suggests that mental stress increases LF activity and decreases HF activity [Williams 1982].

Typical spectral estimation methods such as Fourier transforms typically assume uniform spacing between samples of data and stationarity (i.e., the mean, variance, and autocorrelation structure does not change over time). Uniformly-spaced samples are typically addressed by methods that “fill in” the data in some manner, and non-stationarity issues in the data can be usually be addressed by windowing methods. The Lomb-Scargle method is a well-known methodology, originally developed in astrophysics to detect sinusoidal signals in noisy unevenly sampled times

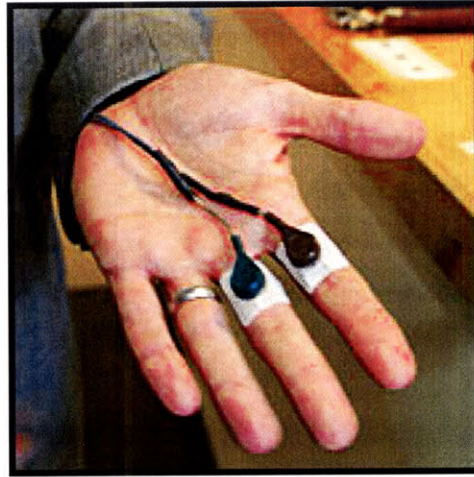
series, which has become a standard tool in the biomedical sciences space to deal with unevenly spaced data [Press & Rybicki 1989]. However, observational data in heart rate can often contain non-Gaussian noise or consist of signals with non-sinusoidal shapes, which can lead to misleading estimates using the Lomb-Scargle methodology [Schimmel 2001]. Qi. et al. has developed a new approach from traditional spectral estimation methods to both of these problems by using a non-stationary Kalman filter within a Bayesian framework to jointly estimate all spectral coefficients instantaneously [Qi et al. 2002]. The technique has been shown to provide more accurate estimation than the Lomb-Scargle method and several classical spectral estimation methods. We used this approach to estimate the power spectrum of the heart rate data to calculate the frequency domain features.

## 4.6 Skin Conductance Response

The heart rate responses discussed above are influenced by both the sympathetic and parasympathetic branches of the autonomic nervous system. In contrast, skin conductance response (SCR), also known as galvanic skin response (GSR) or electrodermal response (EDR), only measures the sympathetic branch. SCR is simply a measure of the electrical conductance of the skin. Simple contact electrodes apply a small voltage across skin, typically measured from the finger and palms. The resulting current that is measured is a function of sweat gland activity and skin pore size. It turns out that the functioning of these eccrine sweat glands, which are concentrated in the hands, feet, and armpits of individuals, are controlled by the sympathetic nervous system. This autonomic nerve response is related to emotional arousal and stress, and can be accurately measured using a simple current detection circuit. As such, skin conductance is one of the fastest responding measures of stress responses. It has been found to be one of the most robust and non-invasive physiological measures of autonomic nervous system activity [Cacioppo & Tassinary 1990]. Selye and others have linked skin conductance response to stress and autonomic nervous system arousal [Selye 1956].

Since skin conductance is well correlated with stress and emotional response, it is a very good baseline sensor for identifying the affective state of an individual. This type of sensing has been well established within the biofeedback and psychotherapy communities for decades. For the LiveNet system, skin conductance data can be obtained with both an armband sensor, or via two

leads from the BioSense board that can be attached to the fingers/palm via cloth-electrodes or finger cuffs.



**Figure 17: Standard skin conductance electrodes placement on the base of the index and middle fingers. Skin conductance is a very useful measure for detecting stress, emotion, and startle responses in individuals based on changes in sweat gland activity controlled by the sympathetic and parasympathetic branches of the autonomic nervous system.**

The skin conductivity response (SCR) sensor detects changes in the electrical conductance of the palm of the hand. The raw skin conductance signal  $S$  can contain high frequency noise, most notably 60 Hz line noise if the LiveNet system utilizes a charger attached to a power outlet. The 60 Hz noise is removed by applying a low-pass filter with a cutoff frequency below 60 Hz (this is not a problem since the skin conductance changes happen on a much lower frequency range). Other random fluctuations can be removed by convolving the original signal with a long averaging window. We convolve the raw signal with a 25-second Hanning window  $h$  ( $s = S * h$ ) to smooth the raw data and characterize the general trend of each affective state.

To account for change in baseline fluctuations (i.e. long-term tonic level shifts) that typically exist in skin conductance measurements, we can use the following normalization technique:

$$g = \frac{s - \min(s)}{\max(s) - \min(s)} \quad 4-19$$

This basically divides the data by the range, resulting in a normal signal that is between 0 and 1, with at least one observed value at each endpoint. This type of normalization technique was proposed by Lyyken, et. al, [Lyyken et al. 1966] and has been used in the psychophysiology field over the years [Lyyken & Venables 1971, Dawson et al. 1990].

SCR features such as normalized mean, standard deviation, and range of the signal can be calculated over various window sizes. One of the most important measures is a measure of the changes in slope of the signal (the positive and negative slopes indicating the sympathetic and parasympathetic activity of the autonomic nervous system, respectively). The instantaneous slope can be effectively characterized using the mean of the first difference of the signal, as follows:

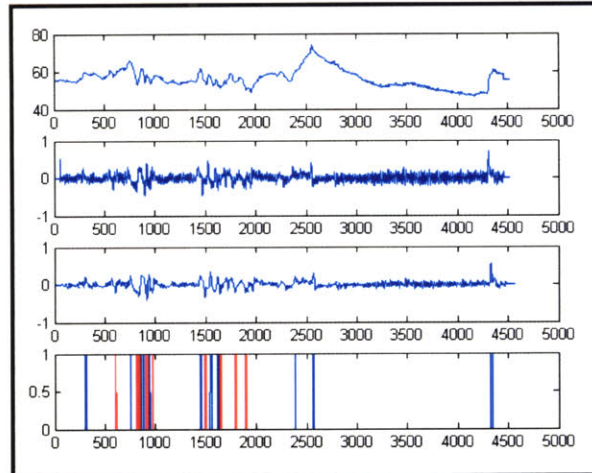
$$g_1 = \frac{1}{N-1} \sum_{n=1}^{N-1} (s_{n+1} - s_n) \quad 4-20$$

Instantaneous spikes in skin conductance are correlated with fear/anxiety/stress/or being startled or other strong emotional physiological responses, the so-called fight or flight response. Longer duration GSR baseline or “tonic” trends, on the other hand, have been related to depression and to longer-term changes in empathy between individuals.

In addition to being a generally accepted measure of mental arousal in people, there have been studies involving looking at collective galvanic skin responses as an objective measure of relationships between married couples or counseling empathy between patients and therapists during psychotherapy [2]. Concordance measures based on correlations in slopes between GSR signals have been shown to be quite good at determining empathy.

By taking a running difference measure of the normalized skin conductance (scaled by the maximum range of the data), we get an approximation for the instantaneous change in slope of the original signal. This signal can then smoothed by applying a low-pass filter, and then thresholded at different levels to get a sense of the number of peaks and valleys in the original skin conductance signal. Small thresholds catch smaller peaks, and larger thresholds can catch very large skin conductance peaks, which turn out to be very correlated to high stress events.





**Figure 18: Raw skin conductance signal trace from the PokerMetrics study showing a response to a number of stressful situations (top), normalized difference (mid top), filtered difference (mid bottom) and negative (red) and positive (blue) thresholds (bottom). The thresholds can be set at different trigger levels to provide a gauge to the number of various-intensity stress or emotion response peaks that occur over a period of time.**

## 4.7 Behavioral Features

In addition to the physiological measures discussed above, the various sensors can be used to detect high-level behavior. A few examples of behavior that can easily be obtained using the given non-invasive sensors are discussed in turn below.

### Sleeping Behavior

Sleeping and diurnal pattern variations can be readily extracted using the data collected in the depression study. A person typically has a routine in their sleep schedule, which can potentially be altered based on factors such as stress and depression. When a person sleeps, they do not move around a lot outside of respiration. Because the armband sensor is located on the back of the arm in a fixed orientation, it is possible to identify posture using the accelerometer by looking at the steady state acceleration effects of gravity. This allows us to determine whether the person is upright or in a lying down position. Sleep can be detected by simply identifying the regions of extended non-movement (falling under a threshold value). However, during sleep a person sporadically shift positions throughout the night. We can identify continuous regions of sleep that take into account shifting movements by considering a sliding window; if the natural shifts in



motion are under a certain window interval, than we can consider the regions of sleep contiguous. Another related behavioral sleep feature that can be extracted is restlessness. Restlessness can be measured by the number of aforementioned shifts in sleep that occur per unit time. Using these algorithms, we can accurately determine sleep duration, restlessness, number of periods of sleep during the day and when they occur as potentially relevant behavioral features of merit.

## **Activity Behavior**

From the accelerometer data, it is relatively straightforward to derive activity, including sustained walking, exercise, or other large-scale body motion sustained over a period of time. This can be measured by calculating motion energy over a certain threshold. In order to remove spurious movements from being considered activity, (such as knocking the LiveNet unit against a hard surface), we can look at the energy over a particular a particular window of time (nominally over a few seconds). In contrast, we can identify sedentary behavior by looking at when the subject is lying down, but not sleeping based on a movement threshold. Activity level and sedentary behavior have both been known to be good indicators of depression.

High-level movement patterns throughout a living environment (amount of time spent in the room vs. out of the room in the common area, for example) is also an indication of social interaction or at least willingness to leave the room. Using the Bluetooth location beacons and the PatientTracker application, we can calculate exactly the amount of time a person spends in a certain area as well as track what their movement patterns are. A number of companies such as Living Independently attempt to do similar people tracking based on motion sensing devices placed around a living space.

## **Social Behavior**

Social interaction is a complex and ubiquitous human behavior involving attitudes, emotions, nonverbal and verbal cues, and cognitive function. Importantly, impairment in social function is a hallmark for nearly every diagnostic category of mental illness, including mood and anxiety disorders as well as dementia, schizophrenia, and substance abuse [DSM 1994]. In addition, social isolation can be a significant stress for patients undergoing rehabilitation from surgical and medical procedures and illnesses. Thus, an important challenge for our behavior modeling

technology is to build computational models that can be used to predict the dynamics of individuals and their social interactions.

Socialization can be measured by proxy by measuring voice activity, as talking is directly proportional to social activity assuming patients do not talk to themselves. Since the speech segmentation algorithm used throughout this thesis can accurately determine the difference between speech and other ambient noises, voice features such as fraction of speaking time can be used to accurately measure socialization levels.

Sensing technologies such as Bluetooth receivers (e.g., those found in modern cell phones) can also be used to quantify social engagement and interaction [Eagle & Pentland 2004]. Similarly, using LiveNet, we can collect data about daily interactions with family, friends, and strangers and quantify information such as how frequent are the interactions, the dynamics of the interactions, and the characteristics of such interactions using simple infrared (IR) sensors and IR tags to identify individuals. Using simple voice features (such as talking/non-talking, voice patterns, and interactive speech dynamics measures) derived from microphones, we can obtain a variety of useful social interaction statistics. We can even model an individual's social network and how that network changes over time by analyzing statistical patterns of these networks as they evolve [Eagle & Pentland 2003, Eagle 2005]. Data on social function can be used as both a marker of improvement or rehabilitation progress or as an indicator of relapse and for use in relapse prevention.

## **4.8 Correlating Non-Invasive Sensing to Clinical Measures**

It is important to establish a correlation between the particular types of cheap, non-invasive sensing that are used by the LiveNet system and more traditionally accepted forms of sensing as a sanity check. The following details some attempts to validate the non-invasive sensors used in this study against more traditional clinical sensing technology.

### **Heart Rate vs. ECG**

Heart rate information obtained from the Polar heart rate chest strap is compared with the ECG designed on the BioSense board (which uses a medical-grade instrumentation amplifier). The

ECG information was filtered through beat detection algorithm and then compared to the binary output of the polar heart rate sensor.

In experiments, the correlation between the Polar heart rate sensor and the ECG was found to be around 0.95 (at rest). The Polar heart rate strap actually performs better than ECG in ambulatory settings because the ECG is particularly sensitive to EMG and motion artifact noise when subjects are moving around. Whereas the ECG can only be used accurately in relative low-motion conditions, the Polar monitor is able to continue to pick up the heart beat information even under considerable movement. This makes sense, as the Polar sensor is designed to be robust even under strenuous exercise conditions such as running and biking.

## **Heart Rate vs. Respiration**

Respiration is an important clinical measure and is correlated to activity state, metabolism, and a variety of other physiological phenomenon. It turns out heart rate variability can also be correlated with other basic physiologic measures, such as respiration. Under resting conditions, healthy individuals exhibit periodic variations in R-R intervals (peaks in an electrocardiogram). This rhythmic variation is known as respiratory sinus arrhythmia (RSA) as it fluctuates in sync with respiration [Cammann & Michel 2002, Hsieh et al. 2003]. Initial analyses indicate that it is possible to accurately estimate respiration rate (a piezoelectric respiration sensor was used as the ground truth for respiration data) from HRV data using the polar heart rate monitor using the LiveNet system.

## **Skin conductance vs. Gold Standard**

Skin conductance via the BioSense board was also tested against gold standard psychotherapy equipment at MGH. In general, the concordance of response between the two systems was very high, with around 80% correlation based on mean value (the only difference between the two systems was magnitude of response).

Arm-based skin conductance via the armband sensor had a markedly more muted response as compared to finger-based GSR measurements. This is consistent with basic understanding of human physiology, since eccrine sweat glands (the type that is affected by the sympathetic nervous system), are much more concentrated in the hands than on the back of the arm where the

armband sensor is placed. Although there is still definitely a positive correlation, it is found that skin conductance results are highly dependent on the individual.

## **4.9 Dealing with Multimodal Sensing and Feature Reduction**

Traditionally, physiological classification and modeling is done by looking in depth at the most dominant physiological feature that is manifested as a symptom of the condition, whether this is movement for Parkinson's disease, ECG patterns for heart conditions, EMG/EEG for epilepsy, etc. While this is the most intuitive and is a great first-approximation method for being able to model and classify various phenomena, it may miss more subtle interactions and correlations in physiological data that may exist that can have diagnostic power.

Thus, in general there is much more diagnostic and predictive power in combining not just traditional physiological measures, but incorporating a variety of simple contextual sensing modalities such as location, activity, and interaction information (at the computer, talking to a person, etc) to aid in classification of high-level human behavior. As an example, a racing heart rate can be indicative of a serious medical heart problem, but with the additional knowledge that the person is rigorously exercising (determined through another sensor such as an accelerometer), the cause of the high heart rate is obvious. Both the context as well as base physiological measures are vital in being able to understand what a person's health state is. These multimodal features are used in the long-term behavior modeling to give more discriminatory power in differentiating between different contexts.

In recent years, a hot topic in ubiquitous computing has been multimodal interfaces but not sensing [Abowd & Mynatt 2000, Oviatt & Cohen 2000]. There have even been a few studies on multimodal physiology-based interfaces that have been initiated [Lisetti & Nasoz 2002]. A goal of this thesis is to discover combinations of physiological and contextual measures that can be used to more effectively differentiate between various relevant phenomena.

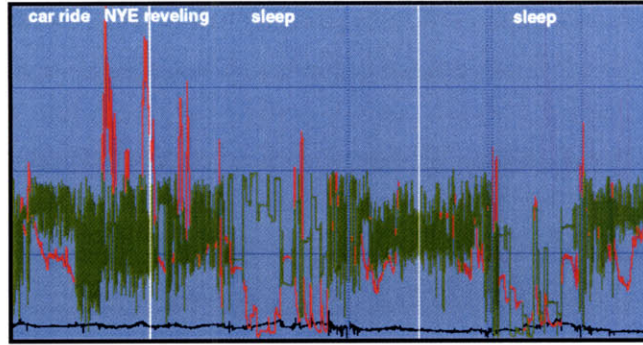


Figure 19: Long-term multimodal heat flux (red), accelerometer (green), and skin conductance (black) data over two typical days of an individual. By simple inspection of the raw data signals, one can visually make out such contexts such as sleeping (periods of low variance in the accelerometer), high activity (during periods with high motion variance and large heat flux spikes), and even when riding in a car (positional acceleration with low heat flux).

Because of the multimodal nature of the physiology data streams that are used in the modeling work, attention needs to be paid to the complications of sensor fusion in the developed models. For example, sensor signals can have different bandwidths or other attributes that may make sensor fusion by forcing into a single framework problematic. Particularly addressing differing bandwidths in sensor streams, it is possible to over-sample/down-sample the slower/faster signal to homogenize data streams, respectively.

## Correlating Features

In order to find the most relevant features for classification and modeling, we can simply run a correlation to determine which features are related to the outcome measures of interest. Correlation is the measure of the relation between two or more variables. Correlation coefficients range from -1 to 1, with -1 representing a perfect negative correlation and 1 a perfect positive correlation. A value of 0 represents a complete lack of correlation. The most widely-used measure of correlation is the Pearson correlation coefficient, also known as the product-moment correlation  $r$ . The Pearson correlation assumes two variables are measured on interval scales (with elements that are not only rank ordered, but the sizes of the differences between elements are also quantified) and determines the extent to which the values of the two variables are proportional to each other. We can calculate the correlation coefficient by summing the products of the standard scores of the two measures and dividing by the number of degrees of freedom:

$$r_{xy} = \frac{\sum_{i=1}^N \tilde{X}_i \tilde{Y}_i}{N-1} \quad 4-21$$

## Principal Component Analysis

Even though there are many potential features that we can use for analysis, it is possible that many of the features are correlated to each other. Thus, more than one feature may reflect the same driving principle governing the behavior or state of the system, and hence there is redundancy in the feature set. Principal component analysis (PCA) is a quantitatively rigorous technique [Strang 1988, Tipping & Bishop 1997] for simplifying a dataset by finding a set of basis vectors that create a subspace that maximizes the variance in the data. In particular, it is a linear transformation methodology that chooses a new coordinate system for the dataset such that the greatest variance by any projection of the data set comes to lie on the primary axis (called the first principal component), the second greatest variance on the second axis, and so on.

Thus, we can reduce the dimensionality of the original dataset while still accounting for as much of the variation in the original data set as possible. Another benefit of performing a PCA is that the observations in the new principal component space are uncorrelated. A method for calculating the principal components uses the empirical covariance matrix of the dataset. We can simply perform an eigen-decomposition to find the eigenvalues and eigenvectors of the covariance matrix.

Starting with the original dataset  $\mathbf{X}$ , we can subtract the sample mean from each row to obtain a data-centered matrix  $\mathbf{X}_c$  that has dimension  $n \times p$ . The sample covariance matrix  $\mathbf{S}$  is simply:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}_c^T \mathbf{X}_c \quad 4-22$$

The  $jk^{th}$  element of  $\mathbf{S}$  is given by

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad j, k = 1, \dots, p \quad 4-23$$

with

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad 4-24$$

We can calculate the eigenvectors and eigenvalues of the matrix  $\mathbf{S}$  by solving the following equation for each  $I_j$ ,  $j, k = 1, \dots, p$

$$|\mathbf{S} - I\mathbf{I}| = 0 \quad 4-25$$

The eigenvectors can be obtained by solving the following set of equations for  $\mathbf{a}_j$ :

$$(\mathbf{S} - I_j\mathbf{I})\mathbf{a}_j = \mathbf{0} \quad j = 1, \dots, p \quad 4-26$$

Any square, symmetric nonsingular matrix can be transformed into a diagonal matrix as follows:

$$\mathbf{L} = \mathbf{A}^T \mathbf{S} \mathbf{A} \quad 4-27$$

Here, the columns of  $\mathbf{A}$  contain the eigenvectors for  $\mathbf{S}$  and  $\mathbf{L}$  is a diagonal matrix with the eigenvalues along the diagonal. We use the eigenvectors of  $\mathbf{S}$  to obtain new variables called the principal components, where the  $j^{\text{th}}$  principal component is given by:

$$\mathbf{z}_j = \mathbf{a}_j^T (\mathbf{x} - \bar{\mathbf{x}}) \quad j, k = 1, \dots, p \quad 4-28$$

The elements of  $\mathbf{a}$  provide the weights or coefficients of the original variables in the new principal coordinate space. The PCA procedure defines a principal axis rotation of the original variables about their means, and the elements of the eigenvector  $\mathbf{A}$  are the direction cosines that relate the two coordinate systems. We can see from the Equation 4-28 above that the principal components are simply linear combinations of the original variables.

We can transform the observations to the principal component coordinate system as follows:

$$\mathbf{Z} = \mathbf{X}_c \mathbf{A}$$

4-29

Here, the matrix  $\mathbf{Z}$  contain the principal component scores, which are the transformed observations, which have zero mean and are uncorrelated to each other. We can scale the eigenvectors to have unit length to produce principal components that have variances that are equal to the corresponding eigenvalue.

Up until this point, we have only performed a transformation, not a dimensionality reduction (the dimensions of the eigenvectors are still  $p$ ). Since the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset, we can project the original data onto the reduced vector subspace formed by the eigenvectors. Since the sum of the variances of the original variables is equal to the sum of the eigenvalues, we can reduce the dimensionality by simply choosing to include the principal components with the highest eigenvalues. In this way, we account for the highest amount of variation with a reduced number of principal components.

An important decision is how many of the principal components to actually use in the analysis. One popular method is to look at the top  $d$  principal components that contribute a cumulative percentage of the total variation in the data, which can be calculated as:

$$t_d = 100 \frac{\sum_{i=1}^d l_i}{\sum_{j=1}^p l_j} \quad 4-30$$

There are various rules of thumb as to how much of the variance is enough; typical advice is to use number of principal components necessary to account for between 70-80% of the variance in the data. Another way is to graphically look at the plot of the eigenvalues sorted by size against the index of the eigenvalue. A typical graph, called a scree plot, is such that the first few eigenvalues are very large relative to the remainder. We can then identify the knee in the plot to identify where most of the variance is explained and choose the corresponding eigenvectors.

We can reconstruct the dataset using the top  $d$  principal components as follows:

$$\mathbf{Z}_d = \mathbf{X}_c \mathbf{A}_d \quad 4-31$$



Here  $\mathbf{A}_d$  contains the first  $d$  eigenvectors of  $A$  and is a  $p \times d$  matrix, and  $\mathbf{Z}_d$  is an  $n \times d$  matrix (each observation now only has the reduced  $d$  elements). The transformation can be inverted to obtain an expression for relating the original variables as a function of the principal components, as follows:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Az} \quad 4-32$$

## 4.10 Model Selection and Classification Techniques

All of the classification and long-term modeling research in this thesis is based on modern statistical machine learning techniques [Duda & Hart 2000, Jain et al. 2000]. These techniques include multiple linear regression, Gaussian mixture models, and hidden Markov models. The mathematical foundations of each of these modeling techniques are outlined in turn below.

### Multiple Linear Regression

The form of a linear model is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad 4-33$$

Where  $\mathbf{y}$  is an  $n \times 1$  vector with the output observations,  $\mathbf{X}$  is an  $n \times p$  matrix with the input observations,  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of the parameters, and  $\boldsymbol{\varepsilon}$  is an  $n \times 1$  vector of uncorrelated noise. Using a least squares fit, the  $\mathbf{b}$  estimates of the unknown parameters  $\boldsymbol{\beta}$  can be found as follows:

$$\mathbf{b} = \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad 4-34$$

By plugging  $\mathbf{b}$  back into the model formula, we obtain  $\hat{\mathbf{y}}$ , the predicted values of  $\mathbf{y}$  given the input:

$$\hat{\mathbf{y}} = \mathbf{Xb} \quad 4-35$$

The residuals are the difference between the observed and predicted  $\mathbf{y}$  values:

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$$

4-36

We can check these residuals for errors in the model, which assumes they have a normal distribution and a constant variance.

## Stepwise Regression

Stepwise regression is an iterative methodology for choosing the input features to a multiple regression model based on the potential for the features to improve the model. There are two ways for the model to be set-up: as a forward or backwards stepwise regression. In a forward stepwise regression, the model starts with zero terms. At each step the most statistically significant term (i.e. the one with the lowest p-value) is added to the model until no more terms can be added that doesn't reach a certain significance threshold (typically the standard  $p < 0.01$  or  $0.05$ ). In a backward stepwise regression, the model starts with all the potential features included, and the least significant terms are removed in turn until the remaining terms are all statistically significant.

The stepwise regression assumes that some of the input terms in the multiple regression do not have an important explanatory effect on the response and hence, it is only necessary to keep the significant terms. A potential problem arises if there is multicollinearity between the input features, where the inputs are highly correlated to each other as with the response. In such a situation, one input term that is included in the model might mask the effect of another input. Thus, the stepwise regression can include different input features depending on the type of model that is chosen and the inclusion criteria.

## Gaussian Mixture Models

Gaussian mixture models [Xu & Jordan 1996] are one of the most widely used types of models in the computational pattern recognition community, given its simple and well-understood assumptions and mature theoretical underpinnings. These generative models can be used both for clustering, density estimation, and classification applications. A general mixture model formulation looks like the following:

$$P(\mathbf{x}|\Theta) = \sum_{i=1}^M \alpha_i h_i(\mathbf{x}|\theta_i) \quad 4-37$$

Where the parameters are  $\Theta = \{\alpha_1, \dots, \alpha_M, \theta_1, \dots, \theta_M\}$ , with  $\alpha_i$  being the mixing coefficients such that  $\sum_{i=1}^M \alpha_i = 1$ , and  $\theta_i$  is the parameterization of the density function  $h_i(\mathbf{x}|\theta_i)$ . For Gaussian mixtures then, the parameterization is simply  $\theta_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$  and the multivariate density function takes the following form:

$$h_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right) \quad 4-38$$

Given a dataset with  $N$  samples drawn from the same Gaussian distribution, i.e.  $\mathbf{X} = \{x_1, \dots, x_N\}$ , and assuming independent and identically distributed (i.i.d.) samples, the resulting probability density function for the samples is:

$$P(\mathbf{X}|\Theta) = \prod_{j=1}^N \Pr(x_j|\Theta) = L(\Theta|\mathbf{X}) \quad 4-39$$

The function  $L(\Theta|\mathbf{X})$  is called the likelihood function of the parameters given the data. The likelihood is a function of the parameters  $\Theta$  where the data  $\mathbf{X}$  is fixed. The problem then becomes maximizing the likelihood function by finding the appropriate parameterization  $\Theta^*$ , as follows:

$$\Theta^* = \arg \max_{\Theta} L(\Theta|\mathbf{X}) \quad 4-40$$

We can maximize the equivalent log-likelihood function,  $\log(L(\Theta|\mathbf{X}))$ , because it is analytically more tractable. For a Gaussian density function, the analysis is tractable and solving for the parameterization can simply proceed by setting the derivative of the log-likelihood to zero and

solving for  $\mu$  and  $\Sigma$ . For other forms, analytical expressions are not possible, so numerical algorithms are necessary.

The EM algorithm is one of the most powerful techniques for finding the maximum-likelihood estimates of the parameters of an underlying distribution from given data set that contains incomplete or missing values [Bilmes 1997]. While the original likelihood function can be analytically intractable, it can be simplified by assuming the existence of additional “missing” parameters. For mixture models, this can be interpreted as the parameter that determines which distribution the data points actually comes from. The EM algorithm creates an iterative procedure to optimize the parameters to maximize the likelihood. This occurs in a two-step process, with an Expectation step (E-step) and the Maximization step (M-step).

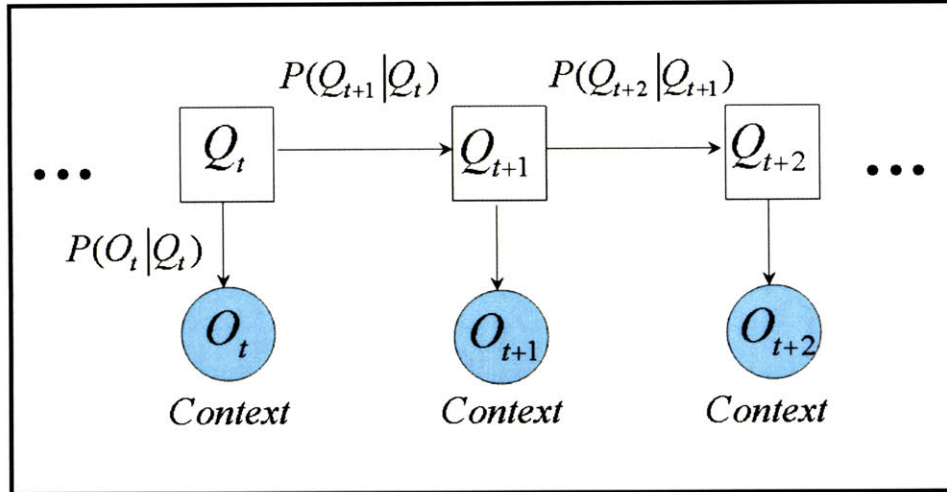
In the E-step, initial guesses for the parameters in the mixture model are made, and each data point is assigned to a distribution to compute the expectation values. In the M-step, the expectation values calculated in the E-step are used to recalculate the distribution parameters. We then repeat the process until there is no further progress made in improving the likelihood. The EM algorithm is a very effective methodology for finding the maximum-likelihood parameters of our Gaussian mixture models, as well as for learning the parameters of the hidden Markov models discussed below.

## **Hidden Markov Models**

Oftentimes, the physiological and behavioral phenomena we are trying to model vary in a time-dependent manner. In order to model the time-dependent dynamics of a system, we turn to Hidden Markov Models (HMMs), statistical models whereby the system to be modeled is assumed to be a Markov process with some hidden parameters which dictate dynamics of the observable outputs [Rabiner 1989].

Hidden Markov models can be viewed as specific instances of a dynamic Bayesian networks (DBNs), directed graphical models that can be used to represent the dependencies between the variables in stochastic processes [Jensen 1996]. In fact, HMMs are one of the simplest kind of DBNs, with one discrete hidden node and one observed node (which may be either discrete or continuous) per time slice. The graphical structure provides a convenient way to specify

conditional independencies in order to provide a compact parameterization of the model. In the following HMM graphical model, squares denote discrete nodes, circles denote discrete nodes, white indicates hidden nodes, and shaded indicates observed nodes.



**Figure 20: Graphical model of a Hidden Markov Model unrolled for 3 time slices. The hidden state (denoted by  $Q$ ), determines the observable output ( $O$ ) which we are trying to model. We can use a given sequence of observations to train a HMM, which can then be used for inference to predict the future behavior of the system.**

The structure and parameterization of the model are assumed to be time-invariant. The HMM parameterization with discrete outputs specifies the number of states, the initial state probabilities, the transition probabilities, output probabilities, as follows:

1. **Number of states,  $N$**  : a scalar indicating the number of hidden states in the model
2. **Number of observation types,  $M$**  : a scalar indicating the number of observation types in the output.
3. **Initial probabilities,  $\Pi = \{\pi_i\}$**  : a vector specifying the probability in each state  $\pi_i = P(q_1 = i)$  for  $1 \leq i \leq N$ .
4. **State transition matrix,  $A = \{a_{ij}\}$**  : a matrix that specifies all the state transition probabilities  $P(q_{t+1} = j | q_t = i)$  for  $1 \leq i, j \leq N$ , where  $q_t$  denotes the current state. These transition

probabilities should follow the normal probability constraints with  $a_{ij} \geq 0$  for  $1 \leq i, j \leq N$  and

$$\sum_{j=1}^N a_{ij} = 1 \text{ for } 1 \leq i \leq N.$$

**5. Output probabilities distribution,  $\mathbf{B} = \{b_j(k)\}$ :** a matrix that specifies all the output emission probabilities  $b_j(k) = P(o_t = v_k | q_t = j)$  for  $1 \leq j \leq N$ ,  $1 \leq k \leq M$ , where  $v_k$  denotes the  $k^{\text{th}}$  observation type, and  $o_t$  the current parameter vector. This matrix should follow probability constraints such that  $b_j(k) \geq 0$  for  $1 \leq j \leq N$ ,  $1 \leq k \leq M$  and  $\sum_{k=1}^M b_j(k) = 1$  for  $1 \leq j \leq N$ .

The resulting formulation for a discrete HMM is thus  $\lambda = \{\mathbf{N}, \mathbf{M}, \mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}$ .

For continuous outputs, we need to specify a continuous probability density function rather than discrete probabilities. We use a Gaussian mixture density function, which is the weighted sum of  $M$  Gaussian distributions. The probability density function is thus parameterized by the mean vectors  $\boldsymbol{\mu}_{jm}$ , covariance matrices  $\boldsymbol{\Sigma}_{jm}$ , and the mixture weighting coefficients  $\mathbf{c}_{jm}$  and can be represented as:

$$\mathbf{b}_j(\mathbf{o}_t) = \sum_{m=1}^M \mathbf{c}_{jm} N(\boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}, \mathbf{o}_t) \quad 4-41$$

Here  $\mathbf{c}_{jm}$  should satisfy the probability constraints  $\mathbf{c}_{jm} \geq 0$  for  $1 \leq j \leq N, 1 \leq m \leq M$  and

$$\sum_{m=1}^M \mathbf{c}_{jm} = 1 \text{ for } 1 \leq j \leq N.$$

The corresponding formulation for a continuous HMM is  $\lambda = \{\mathbf{N}, \mathbf{M}, \mathbf{A}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}, \mathbf{c}_{jm}, \boldsymbol{\pi}\}$ .

Given an observable outputs sequence, there are three canonical problems that can be solved for the HMM:

**1. The Learning Problem:** Given the model  $\lambda$  and output sequence  $\mathbf{O} = o_1, o_2, \dots, o_T$ , how can we adjust the model parameters  $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}$  in order to maximize the probability of observing the observations  $P(\mathbf{O}|\lambda)$ . We can find the most likely set of state transition and output probabilities by using the Baum-Welch algorithm (also known as the forward-backward algorithm) [Bilmes 1997]. The Baum-Welch algorithm is an iterative expectation-maximization (EM) algorithm which computes the maximum likelihood estimates and posterior mode estimates for the parameters based on the observable output sequence as training data.

**2. The Decoding Problem:** Given the model  $\lambda$  and output sequence  $\mathbf{O} = o_1, o_2, \dots, o_T$ , what is the most likely hidden state sequence in the model that produced the observations? We can answer this question using the Viterbi algorithm [Forney 1973]. Given  $N$  states and  $T$  time slices, calculating the probabilities of all transition over time by brute force would require  $N^T$  calculations. The Viterbi algorithm is a dynamic programming algorithm for finding the optimal path. The observation made is that for any state at time  $t$ , there is only one most likely path ending with that state. Thus, when calculating the transitions from this state to states at time  $t+1$ , we don't have to recalculate all the possible paths. Applied to each time step, the Viterbi algorithm can reduce the number of necessary calculations to  $TN^2$ .

**3. The Evaluation Problem:** Given the model  $\lambda$  and output sequence  $\mathbf{O} = o_1, o_2, \dots, o_T$ , what is the probability that a particular observation sequence is generated by the model,  $P(\mathbf{O}|\lambda)$ ? We can evaluate the probability of a particular observation sequence recursively by using the forward algorithm. [Duda & Hart 2000]. The algorithm takes all paths through the model up to the maximum path length  $T$  and calculates the probability of each path, summing them up to give the probability that the model generated the observation sequence. The forward algorithm takes  $TN^2$  steps to calculate this probability.

We have used HMM modeling techniques in order to capture the time dependent dynamics of a number of physiological and behavioral phenomena. Continuous HMMs with Gaussian outputs were used to model the dynamics of shivering in response to cold exposure, as part of the ARIEM soldier monitoring study discussed in Chapter 6. In Chapter 8, we apply a conditional HMM

model to model the long-term behavior of individuals using non-invasive physiology. Specifically speaking to long-term trending, HMMs as a modeling technique has been successfully used to model high-level human behavior in the past [Wren & Pentland 1999, Clarkson 2002, Eagle 2005].



## Chapter 5 Non-Invasive Context Classification

The LiveNet system has proven to be a convenient, adaptable platform for developing real-time monitoring and classification systems using a variety of sensor data. In this chapter, we describe two studies that have been conducted using non-invasive sensing that can identify the high-level contextual state of a person. The first is an accelerometer-based activity-state classifier study, where we show that we can use a minimal amount of sensing (i.e. one accelerometer on the torso) to accurately classify the high level activity and behavior of an individual. The second study attempts to identify stress while players compete in poker tournaments. In this study, we demonstrate that we can use voice, motion, and skin conductance features to create accurate stress and lie-detection classifiers.

Thus, we see that we can use non-invasive sensing both to identify the immediate contextual state of an individual as well as to predict their intent and internal psychological state. Work on these classifiers has been generalized to include critical health conditions, such as a real-time shiver classifier for cold-exposure detection and depression state trending, as described more fully in Chapter 6 and 7 respectively.

### 5.1 General Activity Context Classification

The ability to predict an individual's immediate activity context is one of the most useful sources of contextual information. For example, knowing whether a person is driving, sleeping, or exercising could be useful for a health wearable to calculate general energy expenditure or to initiate a triggered responsive action. Many studies on activity classification have been conducted because of the importance to context-aware systems. Most previous studies on accelerometer-based activity classification that involve multiple activity contexts focus on using multiple accelerometers and require the careful placement of sensors on different parts of the body [Bao 2003, Laerhoven et al. 2002]. A clinical work focusing on using five accelerometers on the appendages and torso has shown that it was possible to detect the onset of symptoms for Parkinson's disease with near 100% accuracy, as compared against physician diary ratings, using a neural network classifier [Klapper 2003]. This study and similar accelerometer-based classification used the LiveNet infrastructure to help develop wearable feedback applications

[Sung et al. 2005]. In another study using a similar setup as the Parkinson's study, it was shown that it is possible to obtain activity classification with an average of 84% for 20 daily activities, with the additional finding that classification accuracy dropped marginally by using only 2 of the sensors (on the wrist and waist) [Bao 2003].

In contrast, we have conducted a study on the use of the minimum set of sensors required to accomplish accurate activity classification. The ultimate goal is to use only a single sensor in random orientation placed close to a person's center of mass (i.e., near waist level), as this replicates the minimum setup requirements of a sensor-enabled mobile phone in the pocket of an individual. The goal is to demonstrate that accurate activity classification can be performed without the need for an extreme level of instrumentation (for example, some systems use up to 30 sensors [Laerhoven et al. 2002] or particular delicacy in the setup (such as the Bao study) in order to achieve good classification results. This way, we can potentially reduce the cost of a recognition system as well as reducing the overall burden when using the technology.

Using a LiveNet system we have been able to discriminate between a set of major activities with classification results in the 81-100% accuracy range using only a single accelerometer located on the torso of an individual [Sung 2004]. The focus of the initial classification study was on common activities such as working at the office, watching TV, and sleeping and transition activities (walking, running, going up/down stairs and elevators). We did not attempt to do any more sophisticated recognition of other activities, though this is entirely within the realm of possibility with the LiveNet framework. Instead, the ultimate goal is to be able to develop a high-level notion of a person's daily activity patterns. In the past, similar motion classifiers to detect head-nodding/shaking in order to identify agreement in classroom settings have been created [Sung & Gips 2005].

## **General Activity Context Experimental Methodology**

For this study, two accelerometer body positions were tested. One was incorporated into an armband sensor, and the other was a waist-worn sensor from the BioSense board in the LiveNet system. The armband was sensor sampled at 32 Hz (maximum frequency of the device), and the waist-worn sensor was sampled at 50 Hz.

This study involved volunteers to perform the required set of activities over the course of a few hours. Data from the following set activities was recorded: working in the office, standing, watching TV, lying down, going up and down stairs, going up and down elevators, riding a car, walking around outdoors, and running outdoors. The total data collection time was around two hours per subject, and involving multiple iterations of repeating each activity. In addition, sleep/activity/sedentary activity data over the course of a week was collected for each subject using an armband monitor and labeled with activity diary information (which recorded changes in contexts and timestamp information). The sample data for the analysis was collected from three subjects who were run through the protocol. It is important to note the data collection occurred in natural settings such as at home and in the office, and was not constrained to just a laboratory setting.

## **General Activity Context Analysis**

From the data collected, mean value, energy, FFT frequency features, and frequency-domain entropy features were calculated using a number of varying sliding time windows (64, 128, 256, 512 samples), corresponding to 1.28, 2.56, 5.12, and 10.24 second intervals. These window sizes all seem appropriate to capture the dynamics of most motion for the activities in question. A standard 50% overlap between windows was used, based on other accelerometer-based studies in the past [Bussmann et al. 2001, Chambers et al. 2002, Foerster et al. 1999].

It turns out that different features were more adept at discriminating between particular activities. For example, mean acceleration values was good for identifying postures (lying down vs. upright) and identifying activities such as riding up and down the elevator, as the effect of gravity can be observed. Energy features were good at differentiating between activities with relatively different energy levels, such as between lying down vs. working at the office/watching T.V. vs. walking or running. Spectral features from FFT transforms performed on the data also yielded useful features (Please see Chapter 4 for more details on the similar modeling methodology used), as different activities can have characteristic frequency attributes. For example, a person might walk at a particular gait, which would result in a peak in that frequency in the spectral features. Spectral entropy can be used to differentiate between activities that might yield similar means and energies. Running, for example, has a much higher spectral entropy than walking, which can be easily observed in a spectrogram.

Combinations of the aforementioned features were used to create a Gaussian mixture model for binary classification for each activity. Each two-state classifier was trained using the training data, and then the leave-one-out cross-validation error was calculated on testing over the entire series of data over all the activities.

## General Activity Context Results

Activity classification accuracies based on single accelerometer features are shown in Table 1 below:

Activity State	Features Used	Accuracy
Standing	Energy, Entropy	95.3%
Walking	Spectral, Entropy	93.4%
Running	Spectral, Entropy	87.7%
Sitting in Office	Posture, Energy	81.3%
Watching TV	Posture, Energy	84.5%
Up/Down Stairs (average)	Spectral, Energy	90.4%
Up/Down Elevator (average)	Posture, Mean	92.1%
Sleep	Posture, Energy	100%
Sedentary Activity	Posture, Energy	100%
General "Activity"	Energy, Entropy	100%

**Table 1: Activity context classification accuracies. Using a combination of spectral features and posture/orientation features using a Gaussian Mixture Model, we can achieve activity recognition accuracies between 81.3-100%.**

The problem areas of misclassification occurred in identifying very similar activities in the same posture, such as between sitting at the office and watching TV. Also, some idiosyncratic movements such as bumping into things or random motions while walking were misclassified as running or walking up and down stairs due to the higher spectral entropy. In general though, we

see that it is possible to very accurately develop binary classifiers that identify single types of activity.

Furthermore, as described in Section 4.7, we were able to develop a number of very useful behavioral classifiers to detect sleep, sedentary behavior, and activity. Based on the accelerometer data collected in the Lifescape Study and basing the ground truth on the diaries information, we have determined that these classifiers are virtually 100% accurate. This is not surprising since sleep/sedentary activity can essentially be determined by positioning data from the accelerometer. Activity can be simply determined and customized based on average accelerometer energy levels passing an adjustable trigger over a specified window of interest. Thus, we can tailor the activity classifier to trigger only on sustained high-level activity as opposed to, for example, bumping the accelerometer or temporarily moving around for a few seconds.

The feature extraction and classification can easily be implemented using the modular spectral feature and Gaussian mixture model subcomponents of the MITHril Real-Time Context Engine. Thus, we can easily port a model developed in Matlab into a real-time classification program that can run locally on a LiveNet system.

## **General Activity Context Discussion**

The above results demonstrate that it is indeed possible to do accurate activity classification using a minimal set of non-invasive accelerometer sensing. This research points to the possibility of identifying a variety of interesting activities of daily living based on very simple accelerometer sensing that does not require multiple sensors on the arms and limbs. While not all activities can be differentiated based on a single accelerometer, this research demonstrates that many interesting activity contexts can be accurately determined based on spectral features and posture. Thus, this research shows promise for being able to develop non-invasive activity context systems that can potentially sewn into the clothing of an individual so as to be completely transparent to the user.

Specifically in the realm of gerontology, being able to determine activities of daily living is key to assessing the status of the elderly, as it has been shown to be a good predictor of nursing home acceptances and mortality [Wiener et al. 1990]. As such, there is a large research effort to identify new technologies that are able to automatically track and classify activities of daily living [Lynch

et al. 2003]. With low-burden sensing such as the type used in this study, it becomes practical to capture a person's long-term activity and health context, which can then be used to detect potentially important deviations from normal healthy behavior.

This research is also important as it indicates the feasibility of doing activity classification using embedded hardware without any specialized setup, wires, or other unwieldy parts. By integrating the accelerometer into an existing device that people are comfortable carrying around (for example, a cell phone), we can significantly lower the bar for developing a practical activity classification system for the mass market that is effectively transparent to the user.

Extensions to the study include conducting a more rigorous analysis of tradeoffs between classification accuracy and sampling frequency. Also of interest would be to incorporate a greedy algorithm feature selection process, to be able to get a sense as to which features are best (as opposed to using the entire feature space or hand-picking the features). Another possibility would be to use a multimodal sensor feature space, where sensor data such as heat flux and heart rate, which is obviously correlated with exertion and activity, could be useful in further improving classification accuracy for more random movements.

In general, we note that only data from three subjects was used to train the models that we have developed, and these individuals were drawn from the same population of graduate students at the Media Laboratory. However, the goal of the study was not to develop a generalized activity classifier, as it is noted that individuals can have very idiosyncratic body movements. Instead, we wish only to show that a particular individual's high-level activity can be accurately classified without a large level of instrumentation. We demonstrate that simple features mean value features, energy, entropy, and spectral features used in conjunction with each other can be very discriminatory.

The activity classifiers developed here are used extensively as behavioral features for our long-term studies. We use a subset of the activity features, the behavioral classifiers (sedentary/activity/and sleep) in Chapter 7 in the MGH Depression Study, to correlate these behaviors with depression state. These behavioral features are also combined together to develop time-dependent models of human behavior at longer timescales in Chapter 8 for the Lifescape Study.

## 5.2 PokerMetrics Stress and Physiology Study

The National Institute of Occupational Safety and Health states that stress is becoming the most prevalent reason for worker disability. A 1992 UN report called job stress "The 20th Century Epidemic", while the World Health Organization stated in 1996 that stress was a "World Wide Epidemic". Researchers estimate job stress costs American industries between \$200 and 300 billion annually. Given the mounting social costs of stress, the possibility of automatically identifying and monitoring stress levels for intervention purposes is compelling. This is particularly true for people who regularly work in high stress environments, such as financial traders or emergency workers. In such situations, faulty performance as a result of acute stress can lead to million dollar losses of or even the loss of life.

Given the implications of stress on work performance, there has been recent interest in monitoring the work performance of individuals under stress. Specifically, studies in behavioral finance present psychophysiological evidence that even the most seasoned securities trader exhibits significant emotional response as measured by elevated levels of skin conductance and cardiovascular variables during certain transient market events [Lo & Repin 2002, Lo et al. 2005]. Other studies have corroborated the evidence linking emotion with trading performance [Steenbarger 2002]. These studies indicate that psychophysiology and stress are intimately linked, and it is possible to infer one from the other.

As a baseline study to determine if it is possible to correlate non-invasive physiology with a variety of important contextual measures such as stress, attention, and interest, a pilot study has been initiated to establish the link between stress and physiology. This study was started to look at player stress within real live poker game scenarios, namely no-limit Texas Hold'em tournaments. In such tournaments, one can lose an entire bankroll in a single hand, termed "going all-in". This has great potential in creating live stressful situations with objective outcomes in a highly controlled setting that can be monitored easily.

We wish to demonstrate that non-invasively derived physiology and behavioral sensing (motion, skin conductance, heart rate, and temperature/heat-flux) can be correlated with various stressful events in poker tournaments. Specifically, we will be looking at subjective reports of stress levels,

bluffing, and all-in situations as outcomes and correlating these situations with physiology features aggregated over the hand.

## **Background on the Physiology of Human Intent and Affect**

Whereas traditional clinical physiology monitoring focuses on identifying accurate physiological responses (e.g. ECG traces during a heart arrhythmia), long-term monitoring enabled by minimally invasive sensing provides the ability to correlate contextual measures over time to a person's behavior and internal state. Research has shown that it is possible to correlate minimally invasive physiology measures to identify notions of a person's intentions and affective state such as interest, happiness, and stress [Picard 2001, Picard et al. 2001].

In particular, accurately identifying human intention such as lying has been particularly contentious given its inherently subjective nature. The concept of the 'lie detector' has always captured the imagination and interest of the popular press since it was invented over a hundred years ago in 1902 by James Mackenzie. The prototypical version invented by Mackenzie, was also called the polygraph tool because it looked at a range of physiological phenomena such as a person's heart rate, breathing rate, blood pressure, and skin conductance while a person is questioned. Though a report released in 2002 by America's National Academy of Sciences indicated that polygraphs are not completely reliable (better as a measure of stress than veracity), there is no doubt that a variety of physiological phenomena can be correlated to lying.

If such a ambiguous thing as telling the truth can be predicted and correlated to basic physiological features such as those analyzed by a polygraph, it is not a large leap of faith to consider using physiology to quantify a person's other internal states such as stress, frustration, interest level, excitement, attention, drowsiness, and affect [Picard 2001, Picard et. al 2001]. In fact, there is research to show that these things can in fact be accurately quantified, and furthermore that these measures can demonstrate a high degree of correlation and concordance among interacting groups of people, whether being stimulated while passively watching a movie [Madan et al. 2004] or the highly involved interaction dynamics of a psychiatric patient and a therapist [Marci 2002].



## **PokerMetrics Human Subjects Approval**

To obtain human subjects approval, we needed to comply with both MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) Department. For this, we needed to provide documentation detailing the experimental protocol, subject consent procedures, and subject recruitment. It was necessary to explicitly describe every type of data collected from the subjects as well the equipment used in the study. Also of particular interest was to make sure the protocol passed the Health Information Privacy Act (HIPA) that protects an individual's right to maintain their health information privacy. In order to comply with these regulations, we had to make sure that personally identifiable subject information was not included in the data that was collected.

## **PokerMetrics Subject Selection**

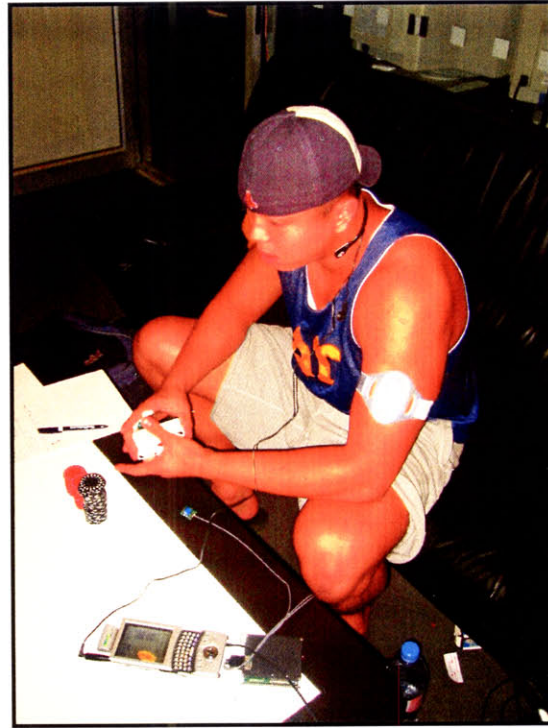
The study population was drawn from the MIT community who answered recruitment emails sent to the MIT poker mailing list. MIT COUHES approval and written consent was obtained from all participants. All of the subjects were college-age, healthy adult male individuals drawn from the undergraduate and graduate schools at MIT. All subjects were self-purported poker aficionados who play semi-regularly, though the skill level ranged from beginners to players who play for profit in high-stakes games.

## **PokerMetrics Experimental Methodology**

For the poker physiometrics study, the physiology and behavioral state of players were taken over 31 tournaments, representing 401 hands of poker. Every subject was paired up with another player to play real live-money games of no-limit Texas Hold'em in heads-up style tournaments. While the players were engaged in the tournament, which typically ranged in duration between 10-30 minutes, the physiology and behavioral state of each player was recorded and annotated using a LiveNet system.

As financial interest was instrumental in generating the stressful situations for this study, the players were asked to play with real money buy-ins in winner-takes-all tournaments. Subjects were paired up to other players with similar financial risk thresholds (with a buy-in between \$5 to

\$20/tournament). The player's self-reported skill level and risk tolerance were also recorded in order to match players as well as to provide additional baseline information.



**Figure 21: Subject wearing the non-invasive sensors (heart rate chest strap, finger skin conductance electrodes, audio microphone, and integrated motion/heat flux/skin conductance monitor along with the LiveNet platform running BioRecord during a session of the PokerMetrics Study. We use the collected physiology to correlate to manually recorded subjective outcomes such as stress level, interest, and bluffing.**

The BioRecord application was used to record multimodal physiology through the duration of each hand, as well as to timestamp interesting annotations during the match. This data included heart rate, skin conductance, temperature, heat flux, speech features, and movement data. The recordings were done on a per hand basis per player. The multimodal data streams of each player were also synchronized to each other.

### **PokerMetrics Subjective and Objective Outcome Data**

In addition to the physiological and behavioral data collected, each subject was asked to fill out a tournament diary at the end of each hand. This form asked information regarding both subjective

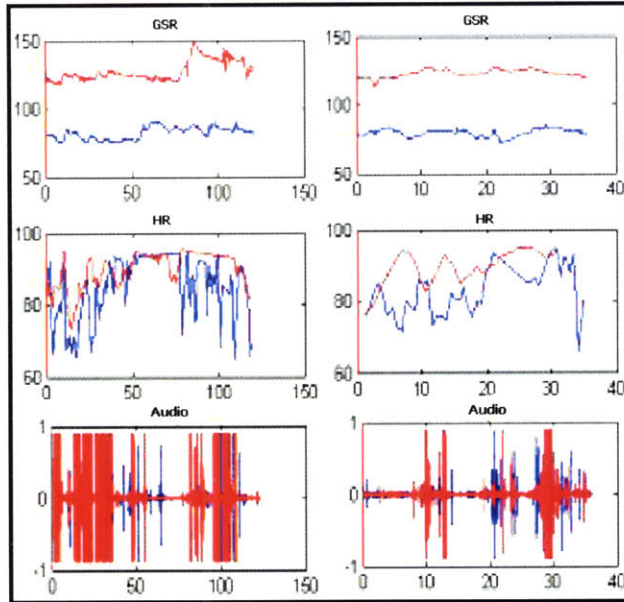
ratings of the player's mental state as well as objective information from the hand. The subjective measures included a player's self-reported interest, excitement, happiness, and stress levels, all rated on a scale of 1-10 (the subjects were asked to rate their current state relative to the typical range of emotional activity that they perceive playing poker). Objective outcome data on whether the individual won or lost, whether they bluffed (i.e. small bluff, medium bluff, large bluff, or semi-bluff), whether there was an all-in hand initiated, whether the hand was a bad-beat (defined as when a heavily mathematically favored hand results in a loss), and the amount of money wagered were also recorded. This combination of subjective measures as well as objective outcomes were correlated with the physiology collected. For the analysis, we focused on the situations in the tournament which are traditionally considered high stress (large bluffing, all-in, bad-beat, and subjective stress situations with a rating of greater than 7 on a scale of 1-10).

## **PokerMetrics Data Analysis**

The physiological and contextual data for each hand of a particular tournament was individually segmented and synchronized. For each hand, a variety of features derived from the heart rate, skin conductance, temperature, audio, and accelerometer data was calculated over the duration of the segment, which typically lasted between 30 seconds for an uneventful hand up to a few minutes for an involved hand.

Given the short duration of each hand, the window for these feature calculations was the entire hand length. Standard heart rate measures such as the average rate, standard deviation in RR intervals, min/max ratios, and frequency-domain energy and entropy ratios were calculated from the hands. For skin conductance, heat flux and temperature, average tonic level, slope variation, and threshold peaks counts were calculated. For the audio features, energy, fundamental frequency, spectral entropy of voiced segments, and variation in these measures over time were calculated, as well as speaking dynamic features such as fractional speaking time, and voicing rates. Motion features such as energy, max/min ratios, entropy, and spectrographic features were calculated.

For each individual's hand, the physiology and behavioral features were computed over the entire hand during the course of play. Two typical hands are shown in the figure below.



**Figure 22: Skin conductance (top), heart rate (middle), and audio (bottom) data for a stressful all-in hand (left) vs. a non-eventful hand (right). As can be seen, the all-in play midway through the stressful hand was anticipated by the blue player who initiates it, shown as a small ramp in the skin conductance trace in the top left graph. This causes a large spike in the skin conductance of the red player a little bit later. We can also see that in general there was increased skin conductance, heart rate variability, and voice activity in the stressful hand relative to the non-stressful hand.**

The hand on the right represents a non-eventful hand synchronized between two players in a tournament. We can see that the skin conductance did not change very much, and there were intermittent spurts of talking. In contrast is the last hand of the tournament between the two same players. The blue player first starts to become stressed (as indicated by the small rise in skin conductivity midway through the hand). A little bit later, he goes all-in, prompting a very large ramp in skin conductance by the red player while he ponders his options. Also note the very large differences in heart rate variability between the two hands (there was much more variability in the stressful hand). Finally, both the magnitude and amount of talking in the stressful hand was noticeably larger in the non-stressful hand.

The features were extracted by partitioning the data of a particular hand into windows (specifically before the resolution of the hand as well as after hand). This was because the voicing and motion dynamics were very different in these regimes since people tend to be more quiet and motionless in stressful situation before hand resolution, and exactly the opposite following the end of the hand.

All calculated features over each hand were normalized and z-scored to find the correlations to the outcomes. There were strong physiological correlations to most of the high-stress situations during play, most notably for all-in plays as well as for large bluffs. For large bluffs in particular, it was found that frequency-based heart variability features such as the LF/HF ratio, fraction speaking time, and motion energy were the most highly correlated features, as shown in Table 2 below:

Bluffing Top Features (n=32)	Statistics	
	R	p
1) HRV (LF/HF ratio)	0.72	0.018
2) Fraction Speaking Time	-0.68	0.021
3) Motion Energy	-0.65	0.03

**Table 2: Top features correlated with bluffing. The heart rate variability LF/HF ratio, fraction of speaking time, and motion energy were the features most highly correlated to bluffing.**

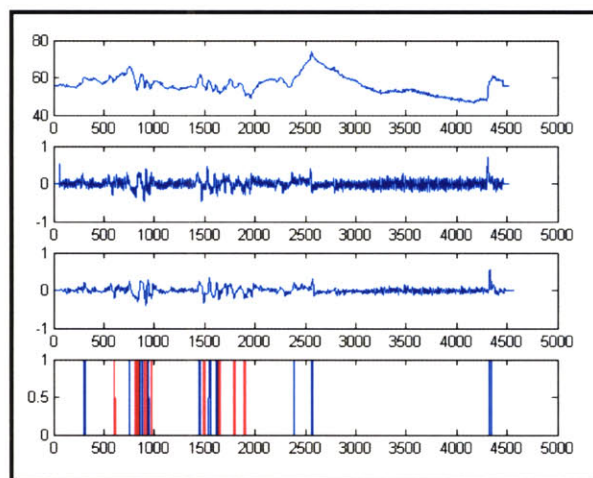
The LF/HF ratio PSD feature was often interpreted as a measure of the relative sympathetic to parasympathetic activity of the autonomic nervous system. The LF power was generated mainly from sympathetic activity, with baroreceptor modulation being a major component in LF power. The HF power, in contrast, was derived from vagal, or parasympathetic activity which is modulated by respiration. Thus, LF/HF ration represents a good indicator of sympathetic-vagal balance, and is used to assess the balance of the autonomic nervous system. As discussed in Section 4.5, studies have shown that mental stress increases LF activity and decreases HF activity. It seems that bluffing, insofar as it induces a stressful situation on a subject, results in a significant increase in LF power, and a slight decrease in HF power. Because of the increase in LF power, the LF/HF ratio was seen to increase during bluffing and other stressful situations.

The fraction of speaking time feature was calculated over the window of the recorded hand with 15 seconds from the end removed (which contains changes in physiology and behavior following the resolution of the hand). This feature was negatively correlated to bluffing.



The third feature, motion energy, was the mean energy calculated over the hand frame and was also negatively correlated to bluffing. We find that there was a significant reduction in the mean motion energy during a hand where the player was bluffing.

The later two features (speaking time and motion energy), are anecdotally well known to be associated with bluffing in the poker community [Caro 2003]. Specifically, when a person was bluffing, he or she consciously tries to reduce all interaction, including to stop speaking and to become very still, so as not to give a reason for the opponent to call the bluff. It is reassuring that this intuition holds true quantitatively in the features that were correlated well with bluffing.



**Figure 23: Raw skin conductance signal trace from the PokerMetrics study showing a response to a number of stressful situations (top), normalized difference (mid top), filtered difference (mid bottom) and negative (red) and positive (blue) thresholds (bottom). The thresholds can be set at different trigger levels to provide a gauge to the number of various-intensity stress or emotion response peaks that occur over a period of time.**

All-in situations, where a player was faced with, or initiated a play where the rest of their money was in jeopardy, was another stress situation. For all-in hands, skin conductance peaks were the most highly correlated features. Below is a figure showing a hand where an all-in play was initiated a little after half-way through the hand. We see that there was a large ramp in skin conductance due to sympathetic arousal from the surprising and stressful situation that has transpired. By taking a running difference measure of the normalized skin conductance (scaled by the maximum range of the data), we get an approximation for the instantaneous change in slope of the original signal. This signal can then be smoothed by applying a low-pass filter, and then

thresholded at different levels to get a sense of the number of peaks and valleys in the original skin conductance signal. Small thresholds catch smaller peaks, and larger thresholds can catch a very large skin conductance peak which turns out to be well correlated to high stress events.

The most basic HRV measure of the standard deviation of RR intervals was the second most correlated features. Voice energy also ended up being the most highly correlated voice feature.

All-Ins Top Features (n=39)	Statistics	
	R	p
1) Skin Conductance Peaks	0.75	0.013
2) Std. Dev. RR Intervals	0.71	0.017
3) Voice Pitch Variation	0.67	0.025

**Table 3: Top features correlated with All-In plays. The number of skin conductance peaks, the std. dev. of the RR intervals of the heart, and voice pitch variation as measured by the std. deviation of the formant frequency were the most correlated features.**

The bad beats results were less correlated to physiology and behavior during the hand, but very correlated to behavior after the hand has transpired. All bad beats had voice and motion energy as well as the standard deviation in the location of the largest auto correlation peak (essentially pitch variation). This makes sense; while the hands varied widely, the defining characteristic of a player being beat was to be very vocal (and whiney about it), as well as to move around in an agitated way following a hand. Statistical significance was an issue since there were relatively few recorded bad beat plays (around 19 instances in 401 hands of poker).

Bad Beats Top Features (n=19)	Statistics	
	R	p
1) Voice Pitch Energy	0.70	0.04
2) Motion Energy	0.63	0.06
3) Voice Pitch Variation	0.54	0.07

**Table 4: Top features correlated with Bad-Beat plays. Voice energy, motion energy, and voice pitch variation as measured by std. deviation of the fundamental formant frequency were the most correlated features.**

Many, but not all of the bluffing/all-in/bad-beat situations described above were highly stressful situations. We define a stressful hand as any in which the player subjectively rated his or her stress to 7 or above (from a scale of 1 to 10). For this measure, a number of the same features described above end up being highly correlated. Skin conductance features, with the number of large-threshold peaks end up being the best feature. It was observed that many hands had small ramps of skin conductance, but large ramps of skin conductance almost always corresponded to highly stressful events. The standard deviation in the largest autocorrelation peak (a measure of pitch variation in the voice) was also highly correlated. The LF/HF ratio shows high sympathetic arousal in the heart rate variability for stressful situations.

Stress Top Features (n=59)	Statistics	
	R	p
1) Skin Conductance Peaks	0.83	0.004
2) Std. Dev. Largest Auto. Peak	0.76	0.011
3) LF/HF Ratio	0.72	0.023

**Table 5: Top features correlated with stress. The number of skin conductance peaks, the std. dev. in the location of the largest auto correlation peak, and the LF/HF ratio heart rate variability measure were the most correlated features.**



## Creating Stress and Lie Detectors from Physiology

The results in the previous section indicate that there were a variety of skin conductance, heart rate variability, voice, and movement features that can be correlated to various stressful outcomes. The question to ask is if these simple physiology features can be used to create accurate classifiers that can discriminate between these outcomes. Of particular interest in a more general setting outside of poker tournaments is stress classification, as well as potentially being able to identify when a person is lying (bluffing).

The leave-one-out cross-validated two-class linear decision classifier for stress has around 79% accuracy using skin conductance peaks. Using two features (skin conductance peaks and voice pitch variation), this accuracy goes up to around 82% ( $\beta_1 = 3.5$ ,  $\beta_2 = 1.9$ ). For the two-feature stress model, the Type-I errors (false positives) and Type-II errors (false negatives) represent 21% and 79% of the total misclassifications, respectively. The classification accuracies for bluffing were a less than those for stress but still significantly above chance. The cross-validated accuracy for a bluffing detector was 64% using skin conductance peaks, and 71% using both skin conductance peaks and voice energy of ( $\beta_1 = 7.3$ ,  $\beta_2 = -4.8$ ). For the two-feature bluffing model, the Type-I errors and Type-II errors represent 38% and 62% of the total misclassifications, respectively.

Classifier	Accuracy	
	Top Feature	2 Features
1) Stress ( $\geq 7/10$ )	79%	82%
2) Bluffing (Large)	64%	71%

**Table 6: Classifier results for stress and bluffing. We can predict high stress events (defined as a subjective rating of 7 for stress on a 10-point scale) with 82% accuracies and bluffing (defined when the subject indicated a large scale bluff) to within 71% accuracies.**

These results seem to conform to the literature in polygraphy (lie-detection techniques) [Podlesny & Raskin 1977]. The various physiological features were well understood to correlate well to stress through the autonomic nervous system reactions. Lying was capable of being detected

insofar as a person becomes stressed at having to lie. In the case of poker, there was obviously a high correlation between the two (around  $R=0.75$ ), as the player was motivated to prevent the potential loss of money. However, for the levels of play in the tournaments that were conducted, it was possible that bluffing becomes something that people adjust to since it was a relatively common occurrence.

## **PokerMetrics Discussion**

These initial results indicate that it is possible to correlate stress, lying (in the context of bluffing), and interest with a variety of physiological features, at least within the context of structured poker settings. Although some of these correlations are no doubt idiosyncratic to the game of poker, it is reasonable to assume that these features may be applicable to more general domains, such as occupational stress. Based on the initial research, we show that it is reasonably easy to develop simple linear classifiers that can accurately predict stress based on short snapshots of physiology.

Using simple linear classifiers, we have been able to identify high stress situations to within about 82% accuracy. We can even detect lying with about 71% accuracy. Essentially, we demonstrate that we can identify these events from simple aggregated physiological features acquired during the duration of the events in question from non-invasively derived sensing. While this leaves room for improvement, the results indicate the possibility of developing classifiers that may be practical in certain domains for triaging people in stressful situations. At the very least, we can create systems that can identify individuals with a relatively high probability of being stressed out, who can then be monitored more closely for potential problems. Specifically within the domain of poker, the ability to identify stress or bluffing even slightly beyond chance provides a compelling advantage.

While the results of the study are very promising, there might have been some limitations to the methodologies we employed to find the appropriate features used in the analysis. Specifically speaking to the HRV measures, they are sometimes derived from windows under less than a minute in duration given the length of the hand (most studies involve between 3-5 minutes of recording due to errors in HRV calculations). For the voice features, many of the poker interchanges are so short in duration (shorts of people saying “call”, “raise”, etc.) that the HMM segmentation algorithm is not able to catch the voicing segments, or construct speaking segments

out of them. Also, the classifiers require features that are computed on the physiology data after the hand is complete, which is not amenable to real-time classifiers which can detect stress or bluffing at the precise moment these situations occur.

The analysis in this study focuses on the outcomes correlated with stress. As we have collected a variety of other tournament information, such as interest levels and objective outcomes like money wagered, it is a potentially interesting exercise to understand how these variables correlate to physiology as well. Another future area of exploration is to use the collected physiology and behavioral data to understand the patterns of arousal and other features to see what precisely is reinforcing addictive behavior and to better understand addiction overall. Gambling in this country is becoming more and more rampant, and in recent years poker in particular has seen a surge in popularity. If we can identify physiological markers for interest, we can potentially begin to develop a physiological model for gambling addiction.

## **Chapter 6 Real-time Clinical Health Classification**

Work on real-time classifiers has also been extended from general context classification to include critical health conditions. An obvious domain for LiveNet is in real-time soldier physiology monitoring and other critical health monitoring applications. For this reason, a research collaboration was started with the Army Research Institute of Environmental Medicine (ARIEM) at Army Natick Labs. The primary objective of the collaboration is to demonstrate the feasibility of using the LiveNet system for these types of real-time critical healthcare applications for the Objective Force Warrior program. The ultimate goal is to be able to develop a variety of real-time wearable health monitors capable of classifying a variety of conditions that soldiers may exhibit through the course of training or combat. As a pilot study, we have used LiveNet systems to collect basic accelerometer and body temperature data of volunteer soldiers who are immersed in cold water for extended periods of time.

### **6.1 ARIEM Cold Exposure/Hypothermia Study**

Army Rangers and other soldiers must perform physically and mentally demanding tasks under challenging environmental conditions ranging from extreme heat to extreme cold. Thermoregulation, or the maintenance of core body temperature within a functional range, is critical to sustained performance [Pozos & Danzl 2003]. Long-term exposure to extreme cold or even moderate cold and wet conditions can result in hypothermia, a condition in which a person's core body temperature drops to a level at which normal muscular and cerebral functions are impaired (clinically defined at less than 35°C). Cold water environments cause body heat loss up to 25-30 times faster than air, so soldiers in water can develop severe forms of hypothermia in very short time periods (survival under extreme cold water exposure can be on the order of hours) [Stock et al. 2004]. In these conditions, it may be advantageous to be able to detect the advent of shivering, as it has been well established that shivering is an initial and key indicator of cold exposure.

In this exploratory study, a real-time wearable monitor was developed using the LiveNet system that is capable of accurately classifying shivering motion through simple accelerometer sensing and statistical machine learning techniques. Based on the collected data, we demonstrate that

shivering can in fact be accurately determined from continuous accelerometer sensing. Results also indicate that specific modes of shivering may correlate with core body temperature regimes, as a person is exposed to cold over time. This exploratory research shows promise of eventually being able to develop robust real-time health monitoring systems capable of classifying cold exposure of soldiers in harsh cold environments with non-invasive sensing and minimal embedded computational resources.

## **6.2 ARIEM Experimental Methodology**

In order to test the feasibility of using the LiveNet system for general soldier monitoring applications, it was decided to first focus on the specific application of identifying and classifying a soldier's level of cold exposure using non-invasive means. Toward this end, a research collaboration was started with the Army Research Institute of Environmental Medicine (ARIEM) at Army Natick Labs. As a pilot study, we have been able to collect basic accelerometer and body temperature data of volunteer soldiers who were immersed in cold water for extended periods of time.

### **ARIEM Study Human Subjects Approval**

Our experimental protocol was piggybacked on a more involved physiological study protocol that the ARIEM facility was running. To obtain human subjects approval, we needed to comply with both MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) Department as well as the ARIEM Internal Review Board. Specifically for COUHES, we needed to provide documentation detailing the experimental protocol, subject consent procedures, and subject recruitment. It was necessary to explicitly describe every type of data collected from the subjects as well the equipment used in the study. Also of particular interest was to make sure the protocol passed the Health Information Privacy Act (HIPA) that protects an individual's right to maintain their health information privacy. In order to comply with these regulations, we had to make sure that personally identifiable subject information was not included in the data that was collected.

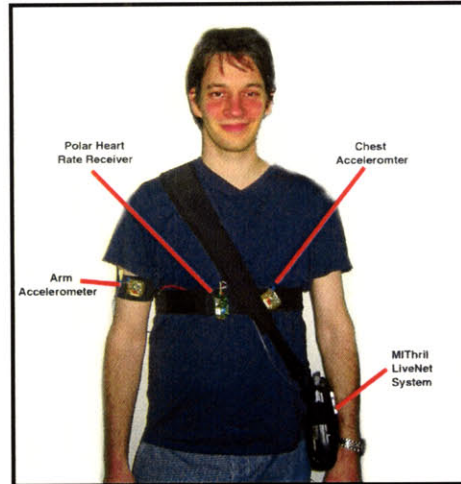
## **ARIEM Study Subject Selection**

The study population was drawn from the U.S. Army Ranger volunteers at the U.S. Army Natick facility. All of the subjects were healthy adult male individuals drawn from the Natick Facility. All subjects were in very good physical and mental shape, and subject to regular army training and conditioning. None of the volunteers had any known serious medical conditions prior to entry into the protocol.

## **ARIEM Protocol**

For this study, we collected accelerometer and core temperature data over the course of thirty experimental sessions between Fall, 2003 through the Fall of 2004, as one part of a larger hypothermia protocol being run by ARIEM. In each session, a volunteer Army Ranger soldier was instrumented with two accelerometers, one on the right arm, and one on the chest, in order to capture the motion dynamics of shivering. In addition, each subject was instrumented with “gold standard” rectal/esophageal thermometers to measure core body temperature.

Each subject was, depending on the larger protocol, submerged between waist and chest deep in cold water (at either 10 or 15°C) and walked at a predetermined speed on a treadmill (which increases water convection, and hence speeds up heat loss). The subject remained submerged until his core body temperature reached the predetermined level of 35.5° C. During this period, the supervising research physiologist continuously monitored the subject’s vital signs. In accordance with IRB guidelines, the subject could stop the protocol at any time, and withdraw from the study for any reason.



**Figure 24: LiveNet system used for the ARIEM study with accelerometer and heart rate sensor network. Given the strenuous operating environment of the protocol, the sensors had to be completely waterproofed and the system shielded to be robust in the presence of large amounts of EM noise. The LiveNet system is capable of not only collecting data, but is also used to prototype the real-time shiver detection classifier that was developed.**

### **6.3 ARIEM Sensing and Data Collection**

The LiveNet system was used both as a data acquisition device to collect the physiology data of the test subject, as well as to implement the real-time shivering monitoring system that was developed.

Motion sensing was accomplished via two embedded microcontroller-based sensors based on the Analog Devices ADXL202 accelerometer part. These devices are two-axis accelerometers (providing acceleration data of  $\pm 2G$  to within 2% precision, in the two dimensions along the plane of the device). Two of these devices are mounted at right angles to create a device capable of measuring 3D acceleration (3 independent axes with one redundant axis). The accelerometer and data was sampled at 50 Hz, time-stamped, and wirelessly streamed to log files on a host laptop.

The LiveNet system proved to be very robust and capable of operating under challenging testing environments. The sensor boards were completely waterproof and submersible, as they had to operate properly for extended periods of time in chlorinated water conditions. In addition, it was important for the systems to be properly shielded in presence of E&M interference. Despite the

noisy RF environment that existed in the ARIEM test facility, the LiveNet system did not have problems wirelessly streaming data to the remote logging machine.

## 6.4 ARIEM Data Analysis and Feature Selection

The core body temperature data over the course of a typical subject run is shown in the figure below. From the graph, we can see three distinct core temperature regimes. The first regime shows the baseline core temperature when the subject was taking mental acuity tests in near room temperature air (in the data, we see that he cools off slightly before stabilizing as he reaches his baseline core temperature at rest). The second regime (labeled ‘Cold’) shows a steady drop in body temperature after the subject was initially immersed into the 15°C water. The third regime (labeled ‘Coldest’) can be identified as a sharper drop in core temperature after prolonged exposure, until the subject reported that he could not continue (about an hour after being immersed).

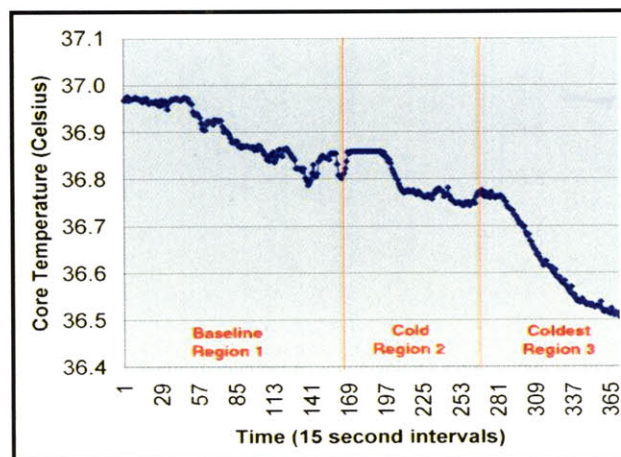


Figure 25: Core temperature regimes of a subject resulting from cold exposure over a period of an hour and a half. We can see that the core temperature of the subject falls slightly as he cools down in room temperature air in the Baseline period (Region 1). Upon entering the chilled water in the Cold period (Region 2), the subject begins to shiver slightly and the core body temperature immediately begins to drop before leveling off at the end of the region. The start of the Coldest period (Region 3) was punctuated with the first signs of intense shivering, and thereafter the core body temperature plummets until the subject indicates he cannot continue with the protocol any longer.

The temperature regimes identified in the graph correlate well with the events of the protocol and the experience reported by subject. Moderate shivering began almost immediately after the

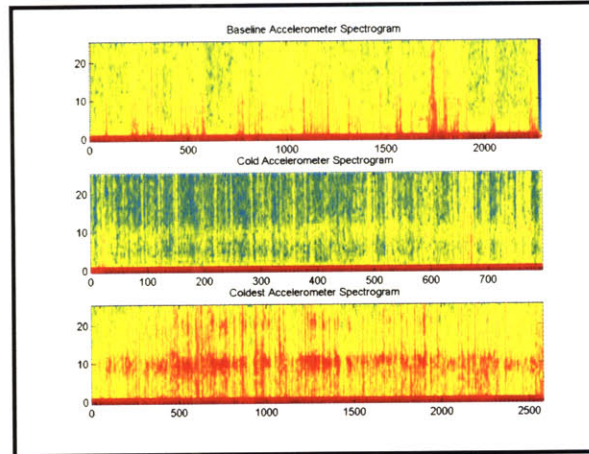


subject was submerged (beginning of the 'Cold' temperature region). At almost exactly the start of the third regime, the subject reported that "it was getting really cold" and the first signs of intense shivering started. The shivering occurred in bursts that visually were very strong and long in duration at first, but over time slowly decayed in both frequency and intensity toward the end of the session.

This observed phenomena is in accordance to literature on hypothermia; shivering can increase heat production, so the body's first response from cold exposure is to start shivering to compensate for the drop in core body temperature. As moderate cold exposure occurs, this shivering can become intense, uncontrollable, and can be episodic in nature, as observed. However, as cold exposure is prolonged, shivering eventually becomes less pronounced and eventually stops in extreme hypothermic states [Pozos & Danzl 2003].

For the classification, the data was divided into the three regions described above: 1) baseline, at rest in air, 2) cold shivering, the first regime following immersion, and 3) coldest shivering, the second regime following immersion. It was discovered that accelerometer data from each of the three temperature regimes identified in the core body temperature data corresponded to distinct features of the accelerometer data. The shivering data had a high frequency signature (around 10 Hz) that was not present in the baseline data, which included activities such as walking and typing on a computer keyboard. This is consistent with a range of shivering frequency between 5-10 Hz reported in literature [Bell et al. 1992], and implies that a frequency-dependent feature is appropriate for classification purposes.

A number of candidate features from the raw accelerometer data were extracted, including binary threshold features, various linear combinations of the magnitudes of three axes of accelerometer data, and power-spectrum features of the raw data. The spectrograms of the magnitude of the accelerometer data yielded the most compelling features, as can be seen visually in the sample raw data in the figure below.



**Figure 26: Accelerometer spectrograms for core temperature regimes identified in Figure 25. We can visually identify the different spectral characteristics for each regime. The Baseline Region spectrogram is characterized by more noise and sporadic episodes of high spectral energy from normal movement as the subject moves from point to point in the facility or bumps into things. The Cold Region spectrogram shows reduced spectral components, primarily due to the dampened movements in water. We see the slight but identifiable 10 Hz component due to the start of light shivering in this region. The Coldest Region spectrogram shows much more pronounced spectral activity, particularly in the 10 Hz region and the 10 Hz harmonics due to shivering. We can use these simple accelerometer spectral features to be able to discriminate between these regions.**

Various power-spectrum features were calculated from an FFT using different-sized windows and overlaps. By comparing the differences between the generated power-spectrums for the different time regimes, parameters that optimally trade off computation time and feature uniqueness were found. The final result that was chosen was a 64-sample FFT window with a 32 sample overlap; this corresponds to a 1.28 second time slice with a 0.64 second overlap between slices.

This particular window-size was chosen as the resulting feature seems to cover enough time duration to capture the dynamics of the various types of shivering. Also, this results in a manageable 32-dimensional feature vector for the model input.

In general, the power-spectrum of the baseline data was noisier than that of the immersed regimes, due to random movement from the subject moving and fidgeting (some of this outlier data was removed to get a more accurate resting baseline). The shivering frequency for the two immersed water regimes occurred in a characteristic band around 10 Hz, with very little power at other frequencies since water immersion causes a dampened motion effect. The shivering data in

the “Coldest” regime also exhibited higher order harmonics at around 20 Hz that was not present in the “Cold” regime.

Although shivering and other motion dynamics were equally discovered on both the arm and chest accelerometer streams, the arm accelerometer data was inherently noisier as the subject often moved his hands around. The chest accelerometer was more stable, and was a better-placed sensor to isolate shivering dynamics. Also, literature on muscle shivering indicates that the magnitude of shivering is higher to centrally located muscles [Bell et al. 1992]. For these reasons, we chose to use the chest accelerometer data as the basis of our classifier development.

We can further reduce the feature space over the calculated 32-dimensional feature vector by simply taking advantage of the fact that shivering primarily occurred around the 10 Hz frequency. Removal of the features that fell well above and below this range did not adversely affect the model accuracy, as they probably do not differentiate among the core temperature regimes and add to the complexity of the model. As such, various feature vector lengths from the power-spectrum were tested with the model. However, it turns out the model performed better with the full feature vector (the model actually needed the higher/lower frequency features to better differentiate between the temperature regimes).

## **6.5 ARIEM Model Selection, Training, and Classification**

Various types of models were considered for a shiver/no-shiver classifier, but it was decided that a simple generative Gaussian mixture model was chosen as a good starting point to be able to capture the structure of the exhibited features.

To facilitate the modeling, the Bayes Net Toolkit (BNT), a set of open-source graphical modeling tools for Matlab developed by Kevin Murphy from the MIT AI Lab was used [Murphy 2001]. Using the BNT, a simple graphical model framework was created. Various Gaussian mixture models were created with different numbers of mixture components and input features. Shivering and baseline data taken from all subject runs was divided into roughly equal training and test data sets. The models were then trained using the training data, with the best shivering model chosen as the one that exhibited the best leave-one-out cross-validation training error. This particular

model was a two-component Gaussian diagonal mixture with a 31- element feature vector (removing the DC spectral component).

The resulting Gaussian mixture models were run on holdout testing data to determine classification error rates. The results indicate that the simple Gaussian mixture model generalized very well to the test data over the experimental runs, resulting in very accurate classification. The classification accuracy rate of the best model on the test data was around 94.7%. It is noted that the accelerometer data is inherently noisy (due to an artifact of the way the sensor reads values, sometimes a sporadic reading is latched to an erroneous value that is railed high, as can be seen in the spectrograms). The only areas of misclassification were around a few areas of episodic noisy baseline accelerometer data where the subject data exhibited a great deal of random motion that happened to mimic the frequency signature of shivering.

In general, the spectrogram windows that were chosen captured the dynamics of shivering and worked rather well. Concerns that the 31-dimensional features used would cause the Gaussian mixture model to overfit and not generalize to the test data were unfounded, as the best model did in fact utilize the full 31-dimension feature vector.

A Gaussian mixture model was also developed to classify core body temperature regime (Baseline/Cold/Coldest) from the accelerometer features. The best model (also a two-component mixture with a 31-dimensional feature vector), resulted in a classification accuracy rate on the test data of around 92%. The model was very accurate in determining core body temperature regime when shivering occurred, but misclassified in regions where shivering stopped during certain areas (more in the Cold region where shivering occurred more intermittently).

Although the Gaussian mixture model worked well to classify the shivering regimes, they fail to capture the complete dynamics of the situation. The exhibited shivering at the beginning of cold exposure first starts as light shivering, then becomes more intense, followed by bursts of shivering, with interval pauses that get longer until shivering finally ceases. A simple Gaussian mixture model trained to detect regimes of cold exposure that only depends on accelerometer spectrographic features would not capture this time-varying behavior. As an example, consider the case of a person already in a state of severe hypothermia such that they were no longer shivering. In such a scenario, the Gaussian mixture model would not take into account the

previous states of shivering and just assume the individual was not shivering, and hence, not hypothermic at all.

To more effectively capture the time dynamics that exist in the shivering data as a person goes through the various stages of cold exposure, a basic continuous HMM was attempted, again using the BNT framework. The model utilized three hidden states and used a 2-component Gaussian mixture.

The resulting HMM was 99.6% accurate on the test data in determining core body temperature regime from the accelerometer data. These results indicate that factoring in the time dynamics of shivering by using an HMM can result in better classification to correlate shivering with cold exposure state than using just a Gaussian mixture model.

## **6.6 ARIEM Real-time Wearable Implementation**

Thus far, all the model implementation was done as a post-processing step in Matlab, after we had collected data. This is an appropriate methodology to perform data analysis and modeling, but is unwieldy as a usable system. It is desirable to find a way to implement the developed model on a wearable embedded system that can actually classify shivering and core-body temperature state from a streaming accelerometer in real-time.

To implement the developed Gaussian mixture model, a collection of streaming data tools and feature extractors was used, called Enchantment and MITHril Real-time Context Engine [DeVaul 2003], respectively. These tools provide a framework to allow one to stream the raw data from the accelerometers through various feature extractors and model implementations.

A real-time FFT feature extractor was used to calculate the spectrogram (in our particular implementation, FFTW, the Fastest Fourier Transform in the West was used as it is very portable and optimized for embedded systems) [Frigo & Johnson 2005]. This feature stream is then passed through a Gaussian mixture model implemented in C. All of this code is optimized such that it can be run on standard StrongARM-based embedded hardware running Linux (we used the Sharp Zaurus SL-5500).

After training the Gaussian mixture model in Matlab, we extracted the relevant model parameters and exported them as a configuration file suitable for use with the MITHril Real-Time Context Engine. (These parameters include the means, priors, mixing weights, and inverse covariances of the Gaussian mixture components.) The Zaurus has enough processing power to perform the FFT and compute the maximum-likelihood Gaussian mixture model classification in real-time.

We evaluated the LiveNet shiver classifier system by streaming the accelerometer holdout data in real-time to the system, and confirmed the same classification performance as on the Matlab model. The system was able to correctly classify most random motion in live testing as non-shivering, as desired. Further testing from more test subjects is required to validate a more generalized classification model.

## **6.7 ARIEM Discussion and Future Directions**

The analysis for the shiver and core temperature classification discussed below was drawn from the first six experimental sessions (the other sessions had not been completed at the time of analysis). We hope to extend this work in the future to be able to generalize the study to see if the results hold true across a broader range of the population.

With regards to core temperature classification, although the trained Gaussian mixture model worked well on the test data for our one available subject run, it is not clear that the model would generalize properly beyond the few test subjects available, as they might have different shivering profiles. In order to generalize to a broader class of individuals, we must also obtain core body temperature profiles from more subject runs and retrain the model with this data. There is a trade-off between model generality and accuracy across different people; more data analysis and modeling is needed to correctly characterize this trade-off.

In addition, the ARIEM protocol also recorded heat flux and heart rate/EKG data that was not used in the current data analysis. Features from these multi-modal sensor streams could be combined to create a more accurate and robust classifier compared to using accelerometer data alone. For example, we were able to perform some initial analysis that indicated that there was a correlation between heart rate variability and shivering intensity.

Initial results of this study are very promising, and indicate that it is possible to accurately classify shiver motion, and that modes of shivering over time can be correlated to core body temperature. This research demonstrates the important fact that we can identify cold-exposure context and core body temperature regimes completely non-invasively using simple accelerometry sensing. Researchers have not yet developed a reliable way of identifying cold-exposure without rather invasive heat sensing. Furthermore, core body temperature, which is the standard way to identify conditions such as hypothermia, has only been possible with extremely invasive sensing (such as device pills or rectal probes). In contrast, we can do an accurate job in detecting cold exposure as well as core body temperature regime using small sensors that can be sewn into the clothing of the soldiers. The classifiers that were developed do not rely on any particular orientation or placement of the accelerometers. In fact, the sensors do not even have to be directly attached to the body of the soldier, which allows us to develop real-time classifier systems that truly do not burden the soldier at all.

The HMM model that was developed did a very good job of triaging core body temperature regime, but was fairly simplistic as it takes advantage of the fact that there is a monotonic decrease in core body temperature in all the subject runs given the nature of the experiment. This may be applicable in a scenario where a soldier is continually subjected to a cold environment. However, in real-world settings, soldiers can have fluctuating body temperatures as they are subjected to different environments. It is not clear that the HMM model would perform as well in a real world environment as a soldier's body temperature ramps up and down, though it is reasonable to assume that motion dynamics would still correlate strongly with core body temperature. In the study, we only correlate shivering dynamics to the core body temperature regime (light shivering for cold regimes, heavy shivering for coldest regimes). In general, we do not see as strong a correlation between the magnitude of shivering and instantaneous core body temperature level, so it is not expected to be able to accurately predict actual core body temperature (as opposed to core body temperature regime) using this methodology.

The ultimate goal of this study is to eventually be able to develop a real-time wearable hypothermia classification system that can determine a person's core temperature state based solely on non-invasive accelerometer sensors. Hopefully, this work can serve as a basis for developing non-invasive monitoring systems that can identify dangerous situations involving cold exposure, potentially saving human lives.

## **Chapter 7      Continuous Monitoring & Clinical Trending**

Thus far, we have used the LiveNet system in studies that have not involved trending of physiology over substantial amounts of time. In both the PokerMetrics Stress Study and the ARIEM Cold-Exposure Study, the monitoring occurred on the order of hours. In addition, in both these cases the studies occurred in constrained environments that did not require a great deal of ambulatory activity (in the PokerMetrics study, the subjects were sitting at a table the entire session, and in the ARIEM study, the soldiers were mostly walking in place on a treadmill in the water for the most of the monitoring). In this thesis, we wish to demonstrate that the LiveNet system is an appropriate platform to be used as a long-term continuous monitoring system in completely ambulatory settings, both in the context of clinical research settings as well as everyday naturalistic settings. We also wish to demonstrate that we can use non-invasive physiology to diagnose, monitor, and trend clinically significant conditions over long periods of time. Toward these ends, a new clinical study in collaboration with the MGH Psychiatry Department is described in this chapter. In this study, the LiveNet system is used to monitor the long-term continuous physiology and behavior of clinically depressed patients, in order to show how non-invasive physiological measures are correlated to depression state as well as to track trends in depression state through the course of treatment.

### **7.1      MGH Depression State and ECT Treatment Study**

Depression ranks among the top health problems worldwide in terms of cost to society. The National Institute of Mental Health states that depressive disorders affect approximately 19 million American adults, almost 10% of the adult population. Depression has been identified by both the World Health Organization (WHO) and the World Bank as the leading cause of disability in the United States and worldwide [Kessler et al. 2003]. Research conducted by the WHO and World Bank shows that major depression is ranked second only to ischemic heart disease in magnitude of disease burden in established market economies [Murray 1996].

Despite the magnitude of this global health problem, there are currently no objective measures based on physiology or behavior that can be used for making a definitive diagnosis for depression, monitoring treatment response, or predicting early signs of relapse. The current



standard for diagnosis is still based on subjective clinical rating scales such as the Hamilton Depression Rating Scale, based on standards that were developed in the early 1960s [Hamilton 1960]. While the efficacy of these scales has been proven in medically diagnosing depression, they have their drawbacks as they are a potential source of subjectivity in the diagnosis as well as the fact that the diagnosis requires the attendance of a physician. Given recent advances in ambulatory monitoring technology such as LiveNet, there is now an opportunity to quantitatively and accurately assess how the long-term physiology and behavior of individuals are correlated to changes in depression.

As part of a collaboration with Dr. Carl Marci from Psychiatry Department at the Massachusetts General Hospital (MGH), a depression study was initiated that combines continuous physiologic and behavioral measures using the LiveNet platform to monitor response to treatment for clinically depressed patients before, during, and after electroconvulsive therapy (ECT). This study, the first of its kind, attempts to correlate basic physiology and behavioral changes with depression and mood state through the long-term, continuous (24-hour) monitoring of clinically depressed patients undergoing ECT.

Electroconvulsive therapy was introduced in the 1930s to treat severe cases of mental illnesses such as severe depression, mania, and schizophrenia. In the treatment, a deliberate seizure is induced in the brain by attaching electrodes to the scalp of the patient and applying short electrical pulses. Today, it is a highly effective form of treatment for last-resort cases where psychopharmaceuticals and other therapies fail [Glass 1985]. Patients who require ECT treatment are a good choice for the study population because of their severe depression. The hypothesis is that sustained chronic depression can potentially induce pronounced changes in physiology and behavior. ECT therapy was chosen due to its high efficacy, relatively rapid onset of effect, and the uniformity of treatment procedures. This allows us to capture the physiology and behavioral changes through the treatment process when the patient responds to the treatment.

The goal of this study is to use continuous ambulatory monitoring technology to develop an effective measure for depression that combines physiologic and behavioral measures consistent with known models of depression and prior clinical research. It is anticipated that changes in these measures (i.e., physiological measures such as skin conductance response, heart rate/heart rate variability, movement, and vocal characteristics and behavioral measures such as activity

context, high-level movement on the ward, diurnal sleep patterns, and socialization) correlate with improvements in clinical rating scales of depression and subjective assessment following ECT treatment for depression throughout the course of hospitalization. These correlations can also potentially be used as predictors or early indicators of clinical response.

There are several aims of this study, the first of which is to demonstrate that we can use the LiveNet platform to create a continuous monitoring system that integrates physiological and behavioral measures appropriate for long-term ambulatory use by clinically depressed subjects. Using the data that is gathered, we are able to test relevant hypotheses for the assessment of response to treatment of major depression in subjects. We show that we are indeed able to use physiologic and behavioral measures to accurately correlate to outcome measures as they change through the treatment process. By taking the most promising features, we have developed simple algorithms that can quantify and classify depression state and effects of ECT based on physiologic and behavioral measures.

## **7.2 Background Depression Physiology Research**

While there are currently no objective physiological markers used in the diagnosis of depression patients, there has been a surge of research activity in recent years that has shed light on both the neurobiological, physiological, and behavioral effects of depression. Modern neuroimaging technologies have advanced our understanding of the neurological effects of depression significantly in recent years [Drevets 2001, Dougherty & Rauch 1997]. Unfortunately, existing neuroimaging technology has several limitations including high cost, difficult maintenance, lack of portability, and the inability to make individual diagnoses [Dougherty & Rauch 2001]. In addition, due to the expense and the side effects of repeated exposure to the neuroimaging technology, data can only be gathered in an episodic manner. As such, neuroimaging techniques are still not practical for use in diagnosing or monitoring depression.

The other promising domain of research has been in studying the physiological effects of depression in clinical research, particularly the types of physiology that can be sensed through the use of non-invasive means. In particular, there have been various studies on correlating depression to measures such as skin conductance response, heart rate, and temperature. Research has shown that the autonomic branch of the central nervous system mediates both skin

conductance response and heart activity [Hugdahl-A 1995]. Changes in skin conductance that are correlated to mental effort, novelty, salience, surprise, and emotionality are mediated through the single innervation of the sympathetic branch of the nervous system [Hugdahl-B 1995]. In contrast, changes in heart rate associated with exercise, mental effort, and emotional states are mediated by the dual innervation of the sympathetic and parasympathetic branches of the nervous system [Hugdahl-C 1995]. Studies have shown that both the sympathetic and parasympathetic response involved with emotional and volitional behavior become less reactive in clinical depression [Buchanan et al. 1985, Oppenheimer & Cechetto 1990, Oppenheimer et al. 1992]. This autonomic blunting, as reflected by decreased sympathetic and parasympathetic tone, can be measured through various physiological cues.

There have been several consistent findings in research on skin conductance response and clinical depression that all suggest a decrease in sympathetic tone in major depression. Decreased skin conductance tone, decreased amplitude of response, and increased non-specific fluctuations have been shown in depressed compared with non-depressed patients [Thorell-A 1987, Thorell-B 1987, Ward et al. 1983]. These findings seem to be consistent across populations from different cultures [Tsai et al. 2003]. Moreover, there is research suggesting that low magnitude skin conductance levels may be a characteristic trait of depressed patients [Thorell-B 1987]. There is also evidence that skin conductance increases with recovery, and patients with recurrent depression and those with a history of attempted suicide often show less increase in skin conductance than in patients that are newly depressed [Thorell & d'Elia 1988].

Research on heart rate and clinical depression has also led to several consistent findings that suggest a decrease in parasympathetic tone regulating heart rate in major depression. In general, patients with depression exhibit an increased baseline heart rate compared with non-depressed subjects [Dawson et al. 1977, Dawson et al. 1985, Lahmeyer & Bellur 1987]. In addition, there have been many studies involving continuous monitoring of heart rate rhythms for psychiatric illnesses, which found that depression patients exhibit a rising heart rate pattern during sleep, explaining an early awakening phenomenon that has been observed in patients [Stampfer 1998, Stampfer et al. 2002, Remick et al. 2005]. Other heart rate studies have shown that normal circadian heart rate rhythms are either reduced or non-existent in depressed patients compared to non-depressed subjects [Taillard et al. 1990, Taillard et al. 1993]. Apathy, a component of depression, has been correlated with reduced HR reactivity [Andersson et al. 1999]. Finally,

consistent findings have shown that patients with chronic depression exhibit decreased heart rate variability, reduced energy in the component bands of the frequency response spectrum, and increases in both in response to treatment of major depression [Chambers & Allen 2002, Carney et al. 1988, Balogh 1993, Volkers et al. 2003].

There is also evidence that metabolism and thermoregulation of the body can be affected by depression. Parasympathetic response by the vagus nerve may indirectly influence thermoregulation by modulating signals that represent information on feeding state, resulting in either reduction or stimulation of metabolic processes [Szekely 2000]. A longitudinal study of circadian body temperature rhythm has shown that increased phase variability, decreased amplitude, and reduced periodicity of circadian rhythm can be correlated to depression and severity of the depressive symptoms [Tsujimoto et al. 1990].

Outside the physiological findings discussed above, there has been a significant amount of literature published regarding qualitative behavioral symptoms of depression that have long been used as indicators for identifying the disorder. Behavioral characteristics such as reduced emotional responsivity, reduced physical activity/psychomotor retardation, decreased socialization/social dysfunction are all well documented as symptoms of depression. In fact, these behavioral symptoms can be viewed as the behavioral manifestations of the autonomic blunting caused by major depression as suggested by the above physiological findings. Each of these behavioral symptoms of depression is discussed in turn.

Psychomotor retardation (defined as the physical slowing down of thought, speech, and movement) is another primary symptom of depression. Psychomotor retardation may be the only group of symptoms of depression that can distinguish depression subtypes and can have a high discriminative validity in determining depression as well as predicting the response of a patient to certain types of treatment such as antidepressants [Sobin & Sackeim 1997]. Motor activity has recently been shown to be reduced in major depression [Volkers et al. 2003]. Research has also shown that there is a correlation between the relative level of physical activity and depression [Camacho et al. 1991]. Because of this correlation, depression is a powerful discriminator between physically active and sedentary individuals [Lobstein et al. 1983].

It is well known that decreased social activity and social dysfunction are symptoms of depression. The quantity and quality of social interaction has been shown to be negatively correlated to depression [Nezlek et al. 1994]. Research has related socialization measures to mental health outcomes such as depression, and that these relations are very similar across ethnic groups [Knight et al. 1994]. In addition, social dysfunction disorders such as social phobia and avoidant personality disorder are oftentimes co-occurrent with major depressive disorders [Alpert et al. 1997].

Changes in diurnal sleep patterns, specifically disturbances of sleep, are typical for most depressed patients and also belong to one of the core symptom groups of the disorder. In fact, there is a strong bi-directional relationship between sleep, sleep alterations, and depression [Riemann et al. 2001]. Studies have shown that it is possible to artificially manipulate the sleep-wake cycle (such as sleep deprivation or a phase advance of the sleep period) which can contribute to mood regulation, alleviate depression symptoms and can even lead to the remission of untreated depression [Cartwright et al. 2003].

While these behavioral symptoms are all highly discriminative in predicting depression, they have traditionally been very qualitative in nature. Using the LiveNet platform, we now have the ability to precisely and quantitatively measure these symptoms to develop objective behavioral measures. For example, by using accelerometers, we can measure the effects of psychomotor retardation on the level of gross body movement. These accelerometers can also be used to identify regions of physical activity, sleep, and sedentary behavioral states throughout the day. We can also use voice analysis techniques to detect the psychomotor retardation effects of depression on speech. Although not specific to the study of depression, speaking patterns as measured by talk time and turn taking have long been used as an index of social interaction [Warner 1992, Watt 1996, Warner et al. 1987]. Thus, by quantifying the speech patterns in the audio data, we can develop an effective measure of social interaction of an individual.

It is proposed that the combination of physiologic and behavioral metrics discussed above would be complimentary in the study of response to treatment in clinical depression. In the following sections, we show that models of depression that integrates behavioral and physiologic measures can be very accurate in determining depression state of a patient as well as trending the effects of treatment.

## 7.3 MGH Experimental Protocol

### MGH Human Subjects Approval

To obtain human subjects approval, we needed to comply with both MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) Department as well as the Partner's Healthcare Internal Review Board at MGH. For this, we needed to provide documentation detailing the experimental protocol, subject consent procedures, and subject recruitment. It was necessary to explicitly describe every type of data collected from the subjects as well the equipment used in the study. The LiveNet monitoring and sensing apparatus was inspected by the MGH Bioengineering Department in order to pass human safety inspection as well as RF communications regulations at the hospital. Also of particular interest was to make sure the protocol passed the Health Information Privacy Act (HIPA) that protects an individual's right to maintain their health information privacy. In order to comply with these regulations, we had to make sure that personally identifiable subject information was not included in the data that was collected. Both COUHES and MGH IRB consent was given for the study by October, 2004.

### MGH Subject Selection

For the study, we are recruiting clinically depressed patients admitted to the MGH Psychiatric Inpatient Unit for ECT treatment to be entered into the study. ECT treatment was chosen due to its relatively rapid onset of effect and the uniformity of treatment procedures. The targeted subject population consists of patients who are diagnosed with major depressive disorder by the DSM-IV criteria [DSM 1994] and have a HAM-D score  $\geq 20$ . Prior inpatient treatment is allowed but not within the last six months. The selection of subjects is limited to English-speaking individuals who are capable of understanding the questions on the questionnaires. Subjects are limited to age 60 or less due to autonomic blunting inherent in normal aging.

To participate in the project, patient subjects must meet the following inclusion and exclusion criteria:

Inclusion Criteria: (1) Age 18-60 (2) Willingness to give written informed consent (3) Meet DSM-IV criteria for Major Depressive Episode (4) HAM-D score  $\geq 20$  upon initial evaluation.

Exclusion Criteria: (1) Unstable medical illness including cardiovascular, hepatic, renal, respiratory, endocrine, or neurologic disorders (including seizure disorder) (2) Substance use disorder active within the last six months or a positive drug screen (3) Psychotic features (current episode or lifetime) as assessed by admitting physician (4) Severe character pathology as assessed by evaluating clinician (5) Currently treated with a medication known to possess anticholinergic properties which is known to interfere with SC measures (e.g., tricyclic antidepressants, clozapine, benztropine) (6) Currently treated with a medication known to possess beta-blocking properties which is known to interfere with HR (e.g. propranolol, atenolol) (7) Clinical or laboratory evidence of hypothyroidism (8) Serious suicide or homicide risk as assessed by evaluating clinician so as to prevent access to monitoring wires (9) Inability to participate safely throughout the study period (3 or more episodes of self-injurious behavior in the past year, documented history of poor treatment adherence) (10) Prior ECT treatment within the last six months.

## **MGH Subject Recruitment**

It is our goal to recruit a total of 20 clinically depressed patients for the study (at the time of writing, a total of six subjects have been run through the protocol). The patient subjects are recruited with the help of MGH Somatic Therapy support staff and psychiatrists to identify potential subjects. Each subject is given a copy of a recruitment flyer at the time of evaluation and consideration as an ECT candidate. Potential subjects are asked if they are interested in receiving a phone call to further explain the study and answer any questions they have prior to admission. Following admission to the MGH Psychiatric Inpatient Unit for ECT, study staff determine the eligibility of the subject in coordination with the admitting psychiatrist. Interested subjects meeting inclusion and exclusion criteria are then given a full copy of the informed consent for further review and contacted by the principal investigator in order to obtain full consent. Written informed consent is obtained from all patient subjects and therapists before protocol-specific procedures are carried out. Subjects are paid \$150 for participation in the study. If a study subject does not complete the full duration of the study, a prorated amount of \$10 per day or \$50 per week is paid.

## **MGH Experimental Procedures**

Once a patient subject agrees to participate in the study by signing informed consent, a full screening medical and psychiatric history is taken by the admitting psychiatrist and a laboratory examination (including thyroid screening and urine toxicology) is performed. Patient subjects are also asked to authorize their therapists to provide the study staff with specific clinical information about current medications and diagnosis. Subjects are asked about current medications, as well as current medical and psychiatric problems. All medication and dosing is recorded, and pre-treatment ratings for the clinical outcome measures is obtained. No other medical information is requested.

ECT therapy is performed at an inpatient unit where the patient stays for the duration of their ECT regiment, which can last anywhere between a few days to many weeks. After an initial evaluation period, a patient is scheduled for ECT therapy treatments three times a week, every Monday/Wednesday/Friday. The total number of ECT sessions is determined by patient response, and can vary between a few sessions to many sessions.

The study requires constant monitoring for the duration of a patient's stay at the hospital during ECT treatment. Every morning, the subject is attended to by study staff to help put on the LiveNet monitoring system, which is worn during all waking hours during the day. This equipment is used to monitor the patient Mondays through Fridays, as well as certain selected days on weekends when staff was available. At night before sleep, the subject removes the monitoring gear, which is stored by nursing staff and replaced by fresh gear in the morning by study staff. An armband monitor is left on the patient 24-hours a day, 7 days a week to be able to monitor and capture diurnal behavioral activity.

At the beginning and end of the hospital stay as well as once a week during the patient's stay at the clinic, patient subjects complete a number of questionnaires administered by an attending psychiatric doctor to obtain clinical outcomes (see descriptions below). Another simple self-reported questionnaire asking about the patient's emotional response is also completed four times a day (before meals and prior to sleep).



## 7.4 MGH Sensing and Data Collection

A LiveNet system is used to provide the ambulatory monitoring equipment necessary to monitor the patient and collect the physiology and behavioral data for the study. The LiveNet system is configured for automated long-term continuous monitoring using the PatientMonitor and PatientTracker applications. The system is worn by the subject to collect data throughout the entire patient's hospital stay.

Every morning, subjects are required to wear a LiveNet system in order to monitor their physiology and behavior during the waking day. The LiveNet system is composed of a small pouch, clipped around the waist like a fanny-pack that serves to record the data, and the associated physiological sensors that are placed on the body. This system is worn continuously while awake and taken off only for the purposes of showering or sleeping. Another armband monitor is also worn continuously throughout the day and night (outside of showering) in order to provide continuous behavioral data. This monitor is worn by the patient 24-hours a day, and even left on during the actual ECT treatment sessions. Thus, we have the full suite of physiology and behavioral data collected from the subjects during the day throughout the course of the hospital stay and throughout ECT treatment as well as 24-hour behavioral monitoring that is useful for diurnal pattern recognition as well as potentially picking up physiology changes during ECT therapy.

In addition, several Bluetooth location beacons are placed in a number of spots within the ward (specifically, the patient's room and in the common areas frequented by patients such as the ward hallway and common dining area). These beacon IDs are recorded by the LiveNet system worn by the patient using the PatientTracker application in order to track the high-level movement patterns of the patients as they moved throughout the ward.



**Figure 27: LiveNet system and associated sensor network used for the MGH Depression Study. In this study, we are measuring body motion, skin conductance (on the fingers and back of the arm), heart rate, heat flux/temperature, and voice activity to find how these types of non-invasive physiology and derived behavioral measures correlate to depression. The patients wear the system continuously for the entire duration of their stay (typically ranging from two weeks onward) for electro-convulsive therapy at the MGH in-patient psychiatric ward.**

## Physiology and Behavioral Measures

Heart rate data is obtained using a non-invasive Polar heart strap worn around the chest underneath the clothing. This sensor can wirelessly transmit heart pulse timing to a receiver board interfaced to the SAK2. The skin conductance sensors are attached to the non-dominant hand of the patient (i.e., the left hand in most subjects) and connected to the BioSense board in order to measure skin conductance response in the fingers. The skin conductance response is sampled at 50 Hz, more than enough resolution to catch the short-term changes in response due to the autonomic nervous system. An accelerometer in the BioSense board is used to measure gross body motion of the patient throughout the day, sampled at 50 Hz. A microphone is clipped to the clothing near the upper chest and interfaced with the LiveNet system. The microphone is used to provide audio feature data, sampled at 8 KHz, to monitor the amount of vocal activity and the characteristics of speech. In order to protect the privacy of the subject, the raw audio data is not recorded so that speech content cannot be extracted. Upper-arm skin response, motion, heat flux,

and temperature data is obtained using the BodyMedia armband sensor, and is sampled at 30 samples per minute to provide coarser resolution behavioral data.

The behavioral measures were largely extracted through the use of the accelerometer data collected. As discussed in the background depression section, activity, sedentary state, sleep, and restlessness have all been shown to be correlated to depression. These behavioral states can all be accurately detected by processing the accelerometer data through the algorithms described in Section 4.7.

## **Clinical Psychiatric Outcome Measures**

In addition to the physiological/contextual measures that are collected, a variety of clinical psychological measures related to depression state, in the form of questionnaires, are collected. These questionnaire forms, specifically the HAM-D, CGI, and QLS described below, are completed by a physician or filled out by the patient upon entering the ECT program, at weekly intervals during their stay on the ward, and the day of release from the hospital. Descriptions of these clinical scales are summarized below:

**Hamilton Depression Rating Scale (HAM-D):** The HAM-D is a widely used, 21-item screening instrument designed to measure the severity of illness in adults diagnosed with depression. It contains items measuring somatic symptoms, insomnia, working capacity, mood, guilt, psychomotor retardation, agitation, and anxiety. The HAM-D has demonstrated validity and reliability for measuring response to treatment [Hamilton 1960].

**Clinical Global Impressions (CGI):** The CGI is a 3-item scale that is used to assess treatment response and is scored by the therapist based upon assessment of the patient's clinical status. The scales are severity of illness, global improvement, and efficacy. Illness severity and improvement are rated on 7-point scales and efficacy is rated on a 4-point scale [Guy 1976].

**Quality of Life Scale (QLS):** The QLS measures patient satisfaction and enjoyment across multiple domains including physical health, mood, work, household duties, school/course work, leisure activity, social relations and general activity [Endicott et al. 1993].

Using these various scales, we have clinically significant outcomes related to a subject's depression and mood state marked at regularly spaced intervals throughout the course of treatment. We can use these clinical outcomes to objectively trend the performance of the treatment and quantify a subject's progress over time. We attempt to correlate physiology and behavior to these clinical outcomes in addition to the subjective outcomes of mood reported by the subject, as described in the following section.

## **Self-Report Emotional Rating Surveys**

In addition to the clinical outcome surveys described above, the experimental protocol also entailed subjects recording their emotional state across a variety of dimensions on a fine-grain basis (4 times daily, after breakfast, lunch, dinner, and right before bed). Toward this end, the special LiveNet survey application Quest was used to create the MGH Emotion Rating Survey (ERS) in order to capture this emotional state. This survey consisted of 26 questions, asking about the subject's current state across a wide variety of emotions. The emotions were chosen in conjunction with an emotion survey specialist for a company that designs emotion-based surveys. The chosen emotions were: 'depressed', 'stressed', 'alert', 'anxious', 'guilty', 'self-worth', 'emotional', 'optimistic', 'energy level', 'hopelessness', 'excited', 'apathetic', 'sociable', 'happy', 'frustrated', 'sleepy', 'angry', 'distracted', 'sad', 'emotionally aroused', 'afraid', 'disgusted', 'bored', 'embarrassed', 'calm', 'tired'.

The ERS survey is based on a visual analog scale via a slider bar that could be moved with a stylus. VAS scales are one of the most frequently used measurement scales in health care research, most commonly known and used for measurement of pain. These scales are used to measure the intensity of sensation and/or subjective feelings [Gift 1989]. The benefit of a VAS survey is that they are intuitive to use, can be administered quickly, and lend themselves to self-completion. To help create appropriate content for the survey, we consulted a company that specializes in PDA-based survey software as well as an emotion rating expert. Computer-based VAS scales have been shown to obtain better resolution of response than hand-scored methods [DeJohn et al. 1992]. In fact, very similar visual analog mood scales were developed for measuring ECT efficacy for cognitively impaired patients, and can be as sensitive to measuring the therapeutic effects of ECT as the HAM-D, as well as being highly correlated to the CGI [Arruda et al. 1996].

The survey questions asked the subject to rate their subjective feelings of intensity to each emotion of interest, with a complete lack of the emotion indexed on the extreme left end of the left end of the line to the highest intensity of that emotion on the right end of the line. For example, for the 'depressed' emotion, we would have "Not depressed" on the left side of the line and "Very depressed" on the right end of the line. This specified a continuous index range of between -100 and 100 for the entire spectrum of the emotion, with 0 being a neutral state that specifies the complete lack of valence in that emotion. We have the precision of a single point, and thus have a resolution over the whole spectrum of 201 index points. In order to provide reference points to help a subject to relatively gauge the intensities on the line, we placed quarter range tick marks above and below the line.

The surveys were designed to be redundant and asked about similar emotional states such as 'depressed' and 'sad' spaced at different points in the survey in order to be able to identify if subjects were consistent with their survey taking. In addition, the survey application recorded the patient's response time, in order to see how long a patient took to answer a specific question. This information was primarily used to detect if a patient was just clicking through questions without bothering to answer them accurately. Any emotion question that took the patient less than 1 second to respond to was not included in the analysis. After the patients became accustomed to the system, they reported that that emotion ratings survey was quite easy to fill out. The ERS survey could be completed in a few minutes.

The end of a survey allowed the patient to voluntarily record any additional relevant information regarding their emotional state, such as if they had visitors, extenuating circumstances, etc. A couple of the subjects provided quite insightful comments (e.g. a patient reporting on one survey that they just got a call from their boyfriend and were in a much happier mood as a result) that could potentially be useful in analyzing and understanding the emotion rating data.

## **7.5 MGH Data Analysis**

A total of six subjects were run through the experimental protocol on the Blake11 Psychiatric Ward at MGH between January, 2005 and May, 2005. Of the six subjects, four of the subject runs consisting of two female and two male patients resulted in relatively complete sets of data through the duration of the ECT treatment stay (the first subject dropped out of the study within

the first couple hours due to an anxiety attack because of the impending first-time ECT procedure, and the second subject run did not yield enough data because the subject exhibited psychotic tendencies and kept on taking off the monitoring apparatus as well as being plagued with some procedural and technical issues). The dataset ranged from one week for the shortest stay, to one patient who stayed over 8 weeks. Average length of stay was 25.8 ( $\pm$  26.2) days. These four data set runs formed the basis of the final data set that is used for analysis described below.

## **MGH Data Preprocessing**

Each full day of patient recording resulted in up to nearly 1 GB of raw uncompressed physiological and audio data along with meta information such as timestamps, data type tags, etc. as well as information from the Bluetooth location beacons and subject-reported emotions rating data. It was common, however, that a day's run did not generate a full gigabyte of data (either because the subject did not put on the monitoring gear until later in the day, or there were technical or procedural issues that prevented the system from recording properly). Also, the PatientMonitor application could encounter a system problem sporadically, and reset the system which starts a new log file. Thus, a day's recording can result in multiple files that can vary in size, timing, and content between days.

The physiology data is first preprocessed by aligning and merging and synchronizing the data files. The metadata is stripped and the data is aligned with timestamps. Each type of physiology data is then individually preprocessed and the all of the relevant features are generated. For all the physiology measures, the z-scored mean, standard deviation, max, max/min ratio, energy, entropy, first and second difference measures were calculated for the skin conductance, motion, heat flux, and temperature data. For the skin conductance response in particular, the data is first smoothed by convolving the signal with a 25-second Hanning window. For the skin conductance response, the standard features were normalized using the baseline normalization technique described by in Equation 4-19.

Since we are only interested in the gross body motion activity of the subject and not in a particular change in orientation, all three channels of the 3D accelerometer data are first merged to create a single magnitude by taking the square root of the sum of the squares of each individual

channel before the features are extracted. The accelerometer data was also used to derive most of the behavioral features by passing them through the behavior classifiers for sleep, restlessness, sedentary behavior, and activity behavior.

There are several data sampling rates used (50 Hz for the skin conductance and motion sensors, 0.5 Hz for the armband sensors, and 8 KHz for the audio). In order to synchronize and make the multimodal data streams uniform, we choose to down-sample the heterogeneous data into a common rate. All of the physiology data streams are uniformly down-sampled to the granularity of one sample per minute by chunking the data using minute-long averaging windows. While the data exists to be able to do more fine grain analysis, such as activity context from the accelerometer data, for our analysis we focus on more coarse-grain analysis that allows us to aggregate the physiology and behavioral data in such a fashion.

Audio data is processed separately. The raw signal is first passed through a number of scripts to generate the appropriate format for the speech segmentation algorithm. Since the segmentation algorithm requires a certain chunk in order to find the voicing and speech segments, we choose to divide the audio data into 10,000 1024-sample chunks (approximately 21.3 minute chunks). This allows the speech segmentation to proceed at a reasonable processing time. The data is also passed through a script to generate all of the 24 voice features described in Section 4.3.

The Bluetooth beacon data is preprocessed by using a script to strip out the location information from the Bluetooth IDs. While the PatientTracker software is primarily used to identify location beacons, it simply does a Bluetooth scan and records nearby enabled Bluetooth devices. As such, it oftentimes recorded the Bluetooth IDs of people carrying mobile phones and other devices. While this information can be used to effectively track the socialization patterns of the patient as well as timestamp interaction with study staff and visitors, for our initial analysis we do not consider this information and strip out any IDs that do not correspond to location beacons. By simply counting the resulting location IDs, we can calculate the relative ratios of the time spent in the room and common areas.

For the subjects run through the protocol thus far, the heart rate data was corrupted due to the fact that the subject staff was not properly trained and placed the Polar receiver too close to a switching regulator on the SAK2 board. Thus, unfortunately, the acquired heart rate data was

spotty and a systematic analysis has not been possible. This problem has been identified and fixed for future subject runs.

In the following analysis section, we first provide the analysis for Subject 6 as a case study of the analytical methodology that is used for every subject. We then summarize the results for each subject, and provide a longitudinal cross-subject analysis in order to identify the results from universal features that are common across subjects.

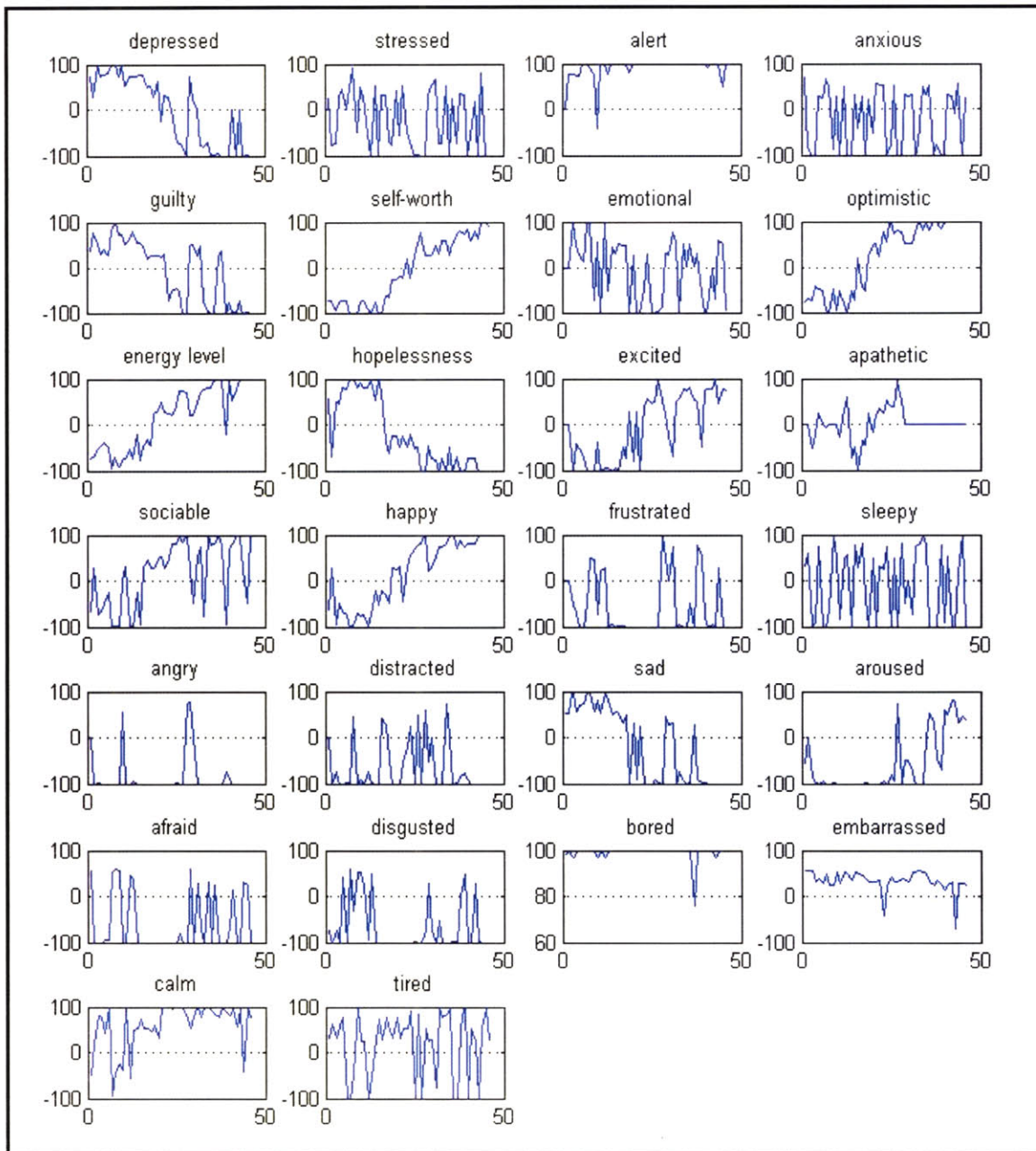
## **Raw Emotional Rating Survey Data**

The emotional rating survey data for the subject is first manually cleaned by removing surveys that were incomplete or where it was obvious that the subject did not fill out the survey appropriately (such as spending less than one second on each question with default value of '0' recorded). The following charts show the self-reported data for Subject 6 through the duration of stay for treatment on the ward.

From the subject's self-reported data, it is apparent that ECT treatment was successful. Looking at the relative trends for correlated outcomes such as depression, guilt, feelings of hopelessness, and sadness, they all significantly decrease approximately midway through the treatment process (in week two of the subject's 3 week stay), while feelings of self-worth, optimism, energy levels, and sociability all increase over the same period.

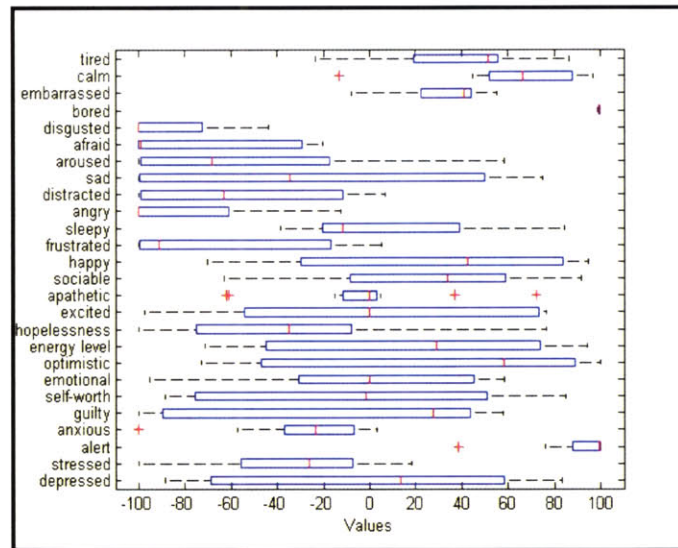
It is observed in the ERS data that particular emotion ratings can fluctuate over a large range due to individual circumstance (effects of visitors, random events, even time of day). Studies have shown that there can be diurnal variations in mood (morning-worse or evening worse patterns) for different types of depression and chronic dysthymia [Rusting & Larsen 1998]. To control for diurnal mood variations as well as to reduce the effects of random fluctuations in emotions, we aggregate the emotion and physiology behavioral sensing data over larger windows to see if we can correlate any features at the granularity of days. Since we are more interested in being able to trend a patient's baseline depression and emotions over long periods of time rather than trying to predict their instantaneous mood at the time of taking a survey, the resolution of a single day is about right.





**Figure 28:** Self-reported emotion rating data for Subject 6 over a 3-week treatment period. Each scale varies from between -100 to 100 on the y-axis, and time progresses to the right. We can see from the ratings that the ECT treatment was successful. Looking at the relative trends for correlated outcomes such as depression, guilt, feelings of hopelessness, and sadness, they all significantly decrease approximately midway through the treatment process (in week two of the subject’s 3 week stay), while feelings of self-worth, optimism, energy levels, and sociability all increase over the same period.

The statistical characteristics of this aggregated emotional rating data (at the granularity of a day) for Subject 6 is summarized below.



**Figure 29: Box and whisker plot of daily-averaged emotion rating data for Subject 6. Boxes have lines at the lower quartile, median (red), and upper quartile values, and whiskers show extent of the rest of the data. Outliers are shown as ‘+’s beyond the ends of the whiskers. We can see that from the diagram that there is a high degree of variation in many of the emotion ratings of interest, specifically those related to depression.**

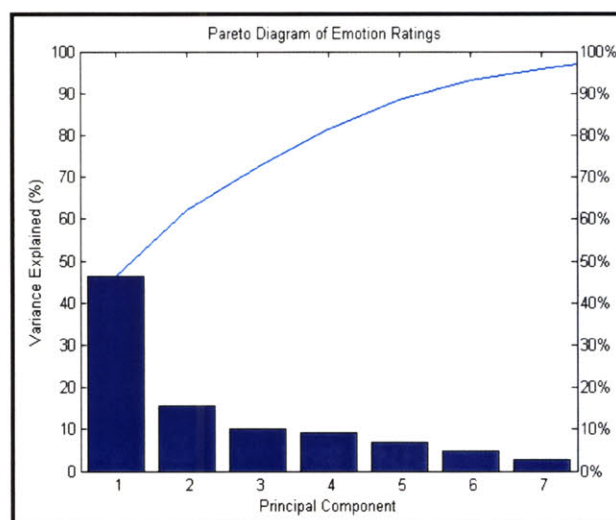
We can see from this diagram that there is a high degree of variation in many of the emotions ratings of interest, specifically those related to depression. This variation in the data may be due to natural fluctuation in the daily moods of the patient, but also is in line with the fact that the treatment process has caused a permanent change in the baseline mood of the patient over time. In contrast, a few emotions ratings, such as ‘bored’ and ‘alert’ for this particular subject do not change very much at all over the treatment, as is expected. A patient could be expected to be suffer from ennui regardless of the treatment, as the ward has very little to offer by way of entertainment. Likewise, for this patient, we see that in general the ECT therapy did not affect the overall alertness of the patient, despite the known short-term grogginess and confusion that may occur immediately following an ECT treatment session.

In general, the emotions of depression, guilt, feelings of hopelessness, sadness, self-worth, optimism, energy levels, and sociability for all the subjects in the study improved for the better through the course of ECT treatment. Many of the other emotions on the ERS survey are also

commonly associated or are co-factors to depression, most notably anxiety, embarrassment, guilt, distractedness, and apathy. The manifestations of these emotions were not universally consistent in all of the subjects, and are more idiosyncratic to the individual circumstances of each patient.

## Decomposing the Emotion Rating Data

Since the emotions used for the ERS survey were picked to capture the various facets of mood related to depression, there is a great deal of redundancy in the emotion ratings data. In order to investigate the underlying structure in the data, we can do a principal component analysis and look at the resulting principal component feature space. We can see the top eigenvalues of the resulting decomposition for the ERS data of Subject 6 in the following Pareto diagram:



**Figure 30: Pareto diagram of the PCA decomposition for the ratings data for Subject 6. From the diagram, we see that the first three principal components explain 46%, 16%, and 10% of the variance respectively, or a total of 72% of the overall variance.**

From the diagram, we see that the first three principal components explain 46%, 16%, and 10% of the variance respectively, or a total of 72% of the overall variance. We can then plot the projected data in the transformed space formed by the first three principal components, as shown in Figure 31.

Each of the 26 emotion ratings is represented in this plot by a vector, and the direction and length of the vector indicates how each variable contributes to the three principal components in the plot.

The first principal component axis can be seen to represent an emotional valence, i.e. positive/negative emotion. All emotion scales related to negative emotions (depressed, hopeless, sad, guilty, embarrassed) are on the positive side of this axis, and all the positive emotions (self-worth, happy, optimistic, energy level, sociable) are on the negative side. The second principal component can be seen to be emotion arousal level or emotionality, with frustrated, disgusted, emotional, stressed on the positive side of the axis and tired, sleepy, calm, distracted, bored, apathetic on the negative side. The third component is a bit more subtle, and can be interpreted as an anxiety axis with the primary scales of afraid/anxious contributing to the negative side of this axis as well as the negative valence and positive emotionality directions of the first two components. This principal component breakdown is consistent with the proposed tripartite model of anxiety/ depression disorder, consisting of 1) low positive affect specific to depression, 2) tension and arousal specific to anxiety, and 3) non-specific general distress, respectively [Clark & Watson 1991, Watson et al. 1995].

Looking at the emotion ratings over the course of treatment, there is a very clear trend to many of them; most of the emotion ratings of interest (including all the positive and negative affect emotions) have baselines that change dramatically over the course of treatment. Outside of these emotion scales related to either positive or negative affect, there is a group of emotions that have to deal with emotional arousal or emotionality (namely stress, anxious, and emotional) that also change with treatment.

The rest of the emotions were a bit more idiosyncratic to the treatment (alert, guilt, apathy, frustrated, angry, disgust, afraid, embarrassed, aroused, distracted, bored, sleepy, calm, and tired) and did not have a clear trend following treatment. In order to remove the ambiguity of how these emotions affect the analysis, we remove these emotions from the PCA analysis to see if it makes the interpretation on the principal components more clear.

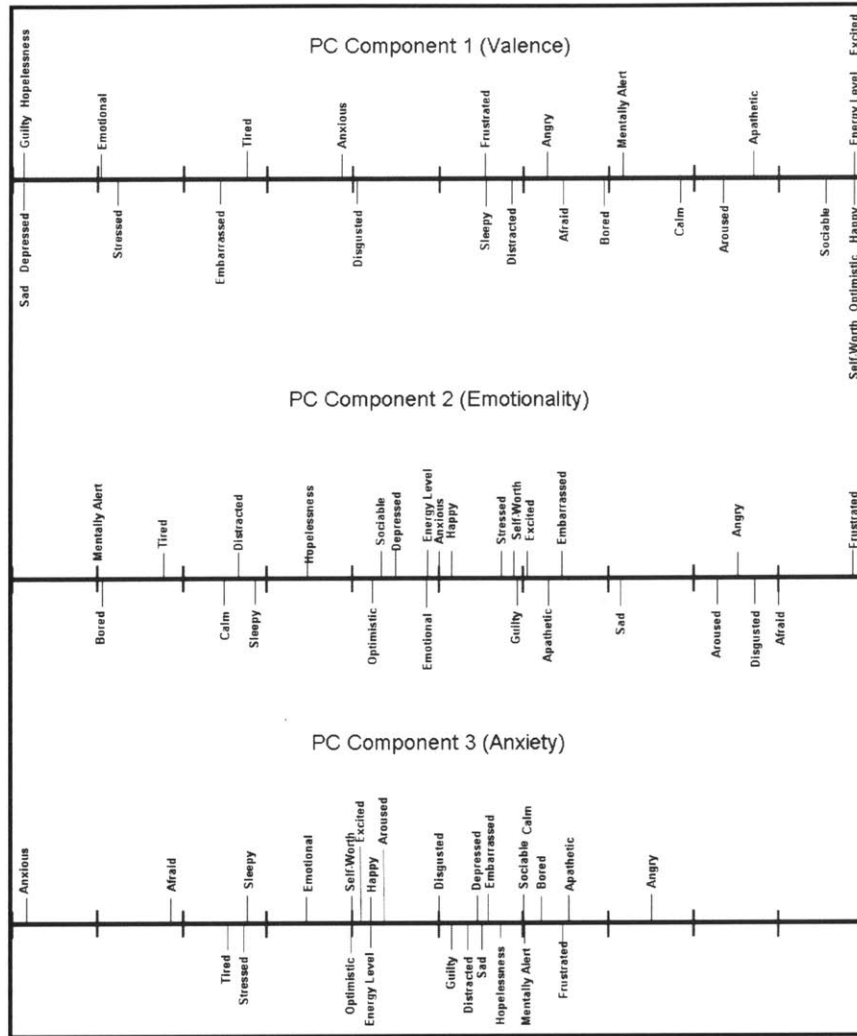
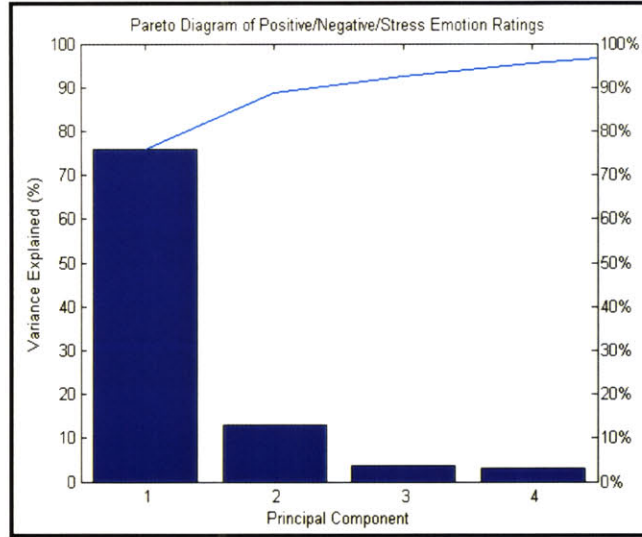


Figure 31: Line charts showing the relative compositions of the first three principal components for the emotion data for Subject 6. PC Component 1 can be interpreted as valence, PC Component 2 as emotionality, and PC Component 3 as anxiety.

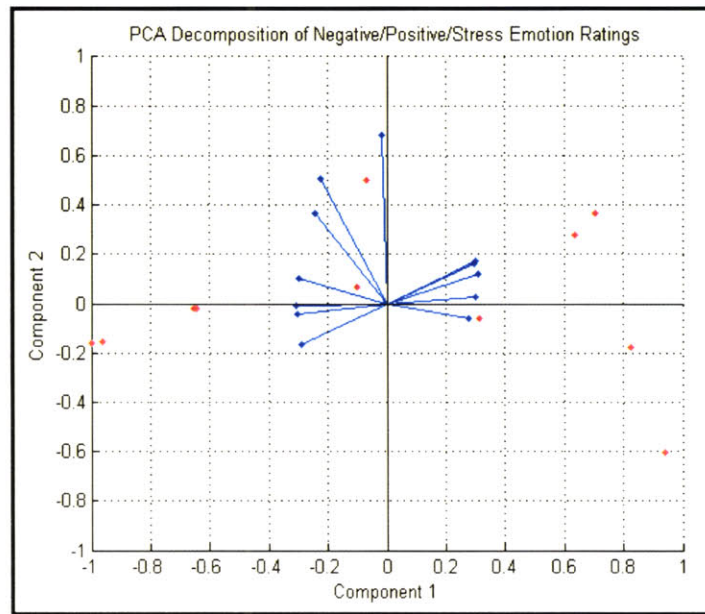
By doing a PCA decomposition of just the depression and emotionality scales, we can see a clear separation of these variables along the principal components. The Pareto diagram show that the variation in the data primarily stems from the first two components, with 76% and 13% of the explained variance respectively.





**Figure 32: Pareto diagram of the valence and emotionality emotion ratings for Subject 6. The diagram show that the variation in the data primarily stems from the first two components, with 76% and 13% of the explained variance respectively.**

The biplot of the scale contributions on the principal components can be seen in the figure below:



**Figure 33: PCA of depression ratings data showing three clusters of negative, positive, and stress emotions. Blue vectors indicate contributions of the emotions to the first two principal components, red dots indicate the data points of the different days during treatment (all days to the left of the y-axis are from the first half of the treatment stay, and all days to the right are from the second half).**

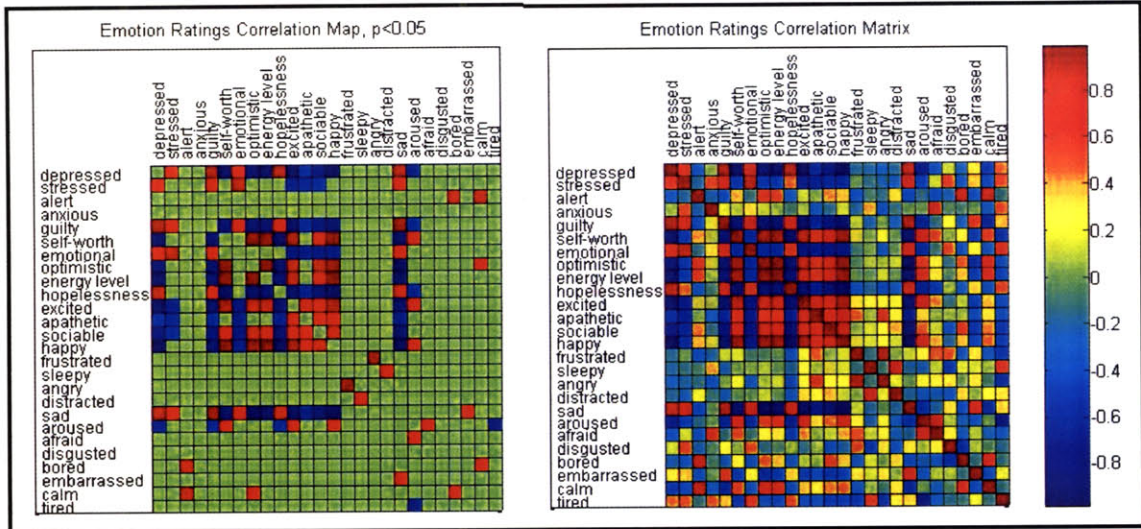
The six vectors directed into the right half of the plot depict the aforementioned positive emotion scales and the four to the lower left of the axis represent the negative emotion contributions. The second principal component, represented by the vertical axis, has positive coefficients for the arousal principal component that exists in the data (stress, anxiety, and emotional scales).

Thus we see that the emotion rating subset of interest can all be effectively reduced into two single dimension components; an emotion rating factor that is exactly the composite emotion rating index derived above, and also a secondary 'anxiety' factor. Anxiety and depression are often highly correlated to each other and a high percentage of the individual disorders often share comorbidities with the other [Brady & Kendall 1992]. In fact, a longitudinal study has shown that where anxiety was present at a moderate severity level, the comorbidity of anxiety and depression is present in 67% of the cases [Zung et al. 1990].

It is important to note all the data points to the left of the y-axis represent the days during the first half of the ECT therapy stay, and all the points to the right represent days during the second half of the stay (that is, treatment was successful in reducing the depression as indicated by migrating to the right of the graph as time progressed). The point at the top middle of the biplot indicates an outlier day (which interestingly has the highest anxiety component of all the days) and corresponds to the unusual spike in the emotion ratings seen across many of the scales of the ERS survey data for Subject 6.

## **Creating a Composite Depression Emotion Factor**

As mentioned above, the emotion ratings survey was designed with redundancy in mind so that we could have more information regarding subjects. As a result, there are a number of similar emotion ratings which are highly-correlated with each other. Looking at the relative trends for correlated outcomes such as depression, guilt, feelings of hopelessness, and sadness, are all highly correlated to each other while being anti-correlated to feelings of self-worth, optimism, energy levels, and sociability over the same period. These emotions are exactly the ones that can be discriminated using the first principal component in the PCA analysis above. The relations between these variables can be seen in the correlation maps below:

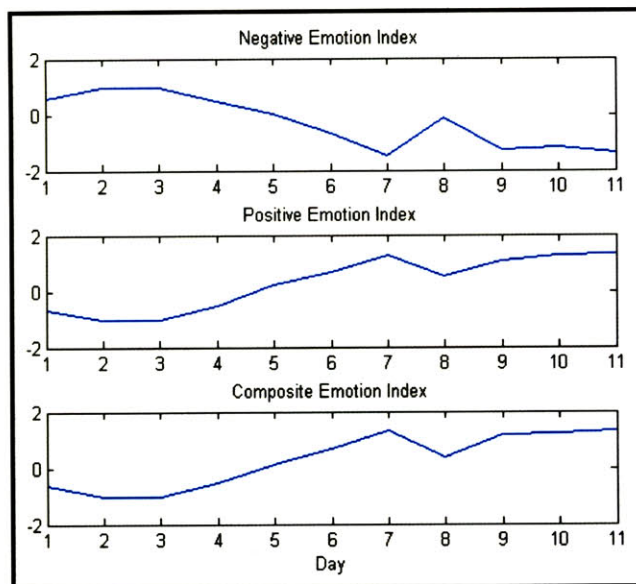


**Figure 34: Cross correlations for  $p < 0.05$  (left) and raw values (right) between emotion ratings for Subject 6. As we can see, many of the emotions such as ‘depressed’, ‘stressed’, ‘guilty’, ‘emotional’, ‘hopelessness’, and ‘sad’ are highly correlated and anti-correlated to feelings of ‘self-worth’, ‘optimistic’, ‘energy level’, ‘excited’, ‘sociable’, and ‘happy’.**

Because the ultimate goal is to create a “depression meter”, we take the highly correlated (and anti-correlated) ratings to depression and merge them together to create a composite index. This can be accomplished by simply creating a linear combination of each emotion rating scale, and serves to reduce the noise in any particular measure as well as combining obviously correlated variables into a single factor that we can predict.

There are three emotion indices that we consider: a negative positive index (composed of the ‘depressed’, ‘guilty’, ‘hopelessness’, and ‘sad’ emotion ratings), a positive emotion index (composed of the ‘self-worth’, ‘optimistic’, ‘energy level’, ‘excited’, ‘sociable’, and ‘happy’ emotion ratings), and a composite emotion index (composed of both the positive and negative emotion indices merged together). Each index is created by adding up the normalized subfactors (scaled to each factor’s range). The composite index is created by adding the positive index to the inverse of the negative index (creating an emotion index that increases as a subject becomes less depressed). These indices for Subject 6 averaged over the days are shown in the figure below:



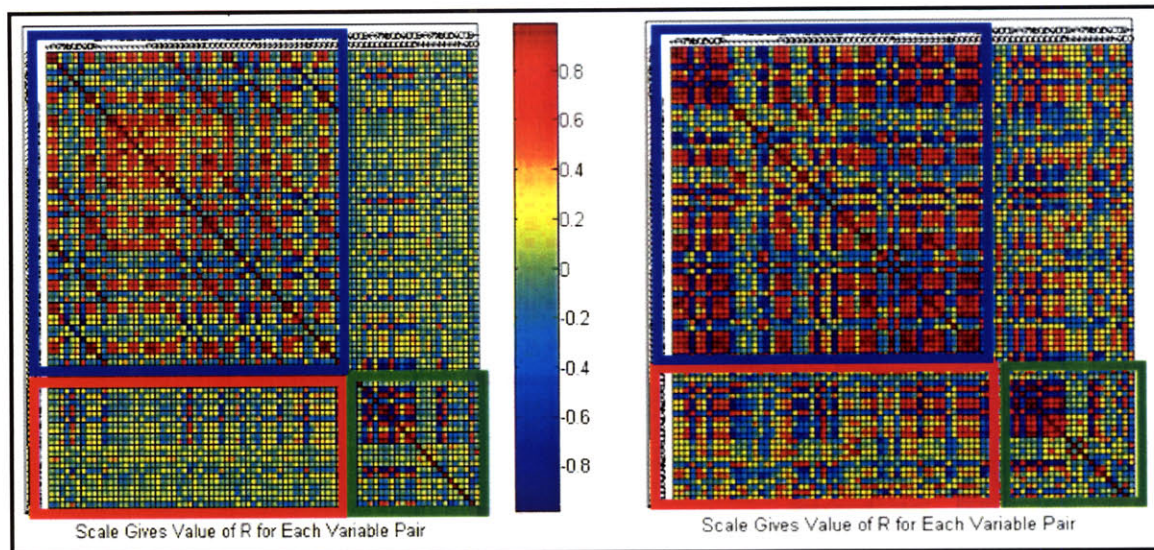


**Figure 35: Plot of normalized daily emotion indices for Subject 6. Note the inverse relationship between the negative and positive emotion indices. The composite emotion index combines the positive emotion index with the inverse of the negative emotion index to create a composite measure with the same trending as the positive emotion index.**

We use these composite depression emotion indices as the responses for our regression analysis with the physiology data discussed below.

## Physiology and Behavioral Data

The physiology and behavioral data is synchronized to the emotion survey timestamps. Using various window sizes (at 5 minutes, 10 minutes, 30 minutes, 1 hour, 2 hour, and 5 hours), the features associated with each type of physiology measure is calculated for every physiology measure centered around the time that each emotion rating survey was filled out. These features include the mean, standard deviation, max, max/min, energy, entropy, first difference, and second difference features calculated for the accelerometer, skin conductance, heart rate, heat flux, and temperature data. For the voice data, the 22 voice features discussed in Section 4.3 are calculated. While there are quite a few correlations over a variety of these physiological features with the emotional rating data, there is a general trend toward larger and more significant correlations the larger the window over which the features was calculated. This can be seen below from the correlations maps depicting different window sizes:

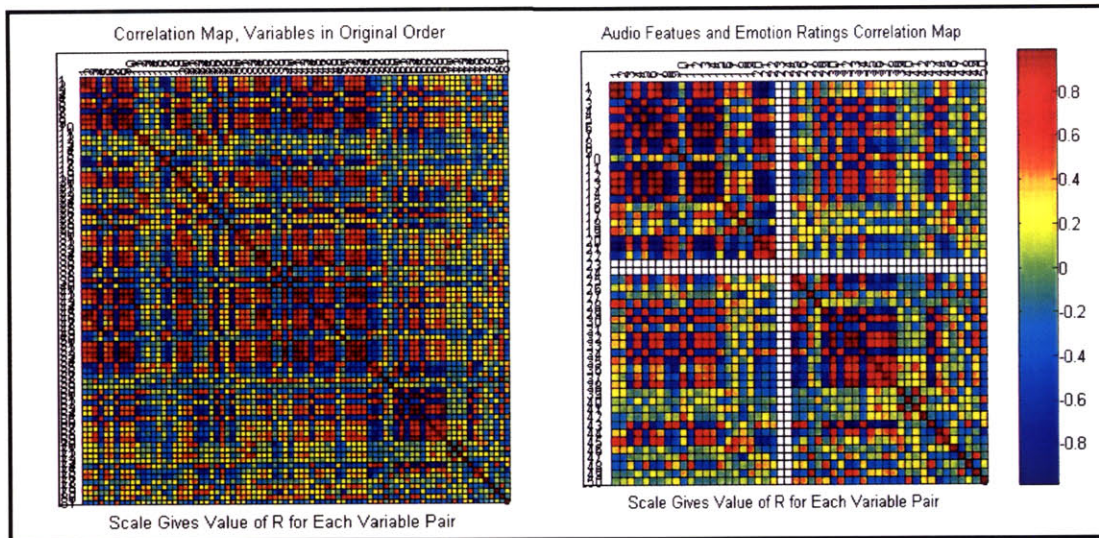


**Figure 36: Correlation maps of raw physiology features and emotional ratings over a 5-minute and 1 day window sizes. The blue box shows correlations between the physiology features, green box shows correlations between the emotional ratings. The red rectangle shows the cross-correlations between the physiology features with the emotional ratings. These correlation maps demonstrate that the cross correlations between the physiology features and emotion ratings become stronger over longer window sizes.**

In order to compare the daily emotion ratings, the physiology and audio speech features are also aggregated over the entire day. Given the fact that the correlation between the emotion data and physiology are stronger for larger window sizes, this strategy seemed reasonable. Both the physiology as well as speech features result in highly significant and strong correlations with the emotion rating outcomes, as shown in the correlation map in Figure 37.

For the physiology features, the strongest correlations are found in the derived features measuring variability in the motion. The standard deviations, maximum value, sum of accumulated differences, energy, and entropy measures for acceleration all show correlations in the range between 0.64-0.81. Likewise, for the audio features, the features involved with the variability of the pitch in the voice, such as the standard deviations in the largest autocorrelation peak, mean formant frequency, and location of the autocorrelation peak have correlations between 0.69-0.82.





**Figure 37: Aggregated physiology (left) and audio (right) features and emotion rating correlations for Subject 6 at the granularity of a day. Notice that these aggregated correlations are much more highly pronounced than in Figure 36. We use step-wise regression techniques to find the most correlated features to use for regression analysis with the subjective and clinical outcomes.**

In addition to the physiology and voice features discussed above, the accelerometer data from the armband monitor is passed through the activity context, sedentary behavior, sleep, and restlessness detection algorithms discussed in Section 4.7 to generate the activity context behavior features. These behavior features are binary in nature (true/false), so just the basic statistical features (mean, standard deviation, and on/off ratio) through the course of the day are generated. The exception is the restlessness feature detector, which essentially outputs the number of times a subject flips while sleeping). These accelerometer-based behavioral measures and the location beacon feature (in-room vs. out-of-room time ratio) form the basis for our behavioral features used in the study.

In the following sections, we use these aggregated physiology and behavioral features for our regression analysis with the composite emotion indices developed above.

## Depression Regression Analysis

Looking at the physiology and speech features, it is apparent that there are a multitude of potential features at the  $p < 0.05$  significance level. However, at that significance level, one out of every twenty features would be statistically significant by chance, and we have quite a few

candidate features to choose from. It is reasonable to assume that some of these features do not have an important explanatory effect on the emotional outcomes, and hence can be removed to simplify the model by reducing it to only including the most significant features necessary to explain the response.

Thus, we employ a stepwise multi-linear regression analysis to choose the most appropriate features to include in the multiple regression model. Two strategies are employed:

**Forward Stepwise Regression:** A regression model which starts without any model terms, and adds the most statistically significant feature (the one with the lowest p-value) at each step until the stop criterion is reached.

**Backward Stepwise Regression:** A regression model which starts with all the features in the model and removes the least significant features until all the remaining features that exist in the model are statistically significant.

First we perform a stepwise regression using just the physiology and voice features as candidate inputs and the composite emotion factor as the output. It turns out that both the forward and backward stepwise regressions results in the selection of the same features for Subject 6. The top three chosen features are shown in Table 7 below:

Multivariate Regression Emotional Factor Top Features	Forward Stepwise Regression		Backward Stepwise Regression	
	R <sup>2</sup>	p	R <sup>2</sup>	p
1) Std. Dev. of the Location of the Largest Auto Correlation Peak	0.68	0.0011	0.68	0.0087
2) Max Value Acceleration	0.65	0.014	0.65	0.0056
3) Max SAD Acceleration	0.64	0.018	0.64	0.014

**Table 7: Top features selected for multivariate regression against composite emotion index (chosen with a p<0.05 criterion) for Subject 6. Std. dev. of the location of the largest autocorrelation peak, maximum value in acceleration, and maximum in the sum of accumulated differences were the top three performing features.**

The above features can be interpreted as the 1) variance in the frequency response, or pitch, of the voice, 2) the relative maximum in magnitude of the motion, 3) the magnitude of the variance in the motion. Essentially, the model can be interpreted as saying that the pitch variation in the voice and the relative amount of motion of the subjects as aggregated over the day is the best predictors of the emotion ratings we are interested in.

One potential problem of multiple regression analysis is the multi-collinearity of related features, in which case the presence of one feature could mask the effect of another feature. We can see in the following correlation matrix how these chosen features vary with the other audio features. However, we note that the same set of features is chosen using both forward and backward regression. It is comforting that the choice in the stepwise selection process does not seem to affect the features selected to for the regression.

Using the top two features in Table 7, we can now calculate the linear regression results for the various emotion ratings as well as depression emotion indices. These results are summarized along with their  $R^2$  and  $\beta$  values in Table 8 below.

It is interesting to note that we can see that the positive emotion ratings (self worth, optimistic, energy level, excited, sociable, happy, emotionally aroused) all have large negative slopes for both the first and second features, and negative emotion ratings (depressed, guilty, hopelessness, and sad) have large positive slopes. The  $R^2$  of all the emotion indices are between 0.79-0.85.

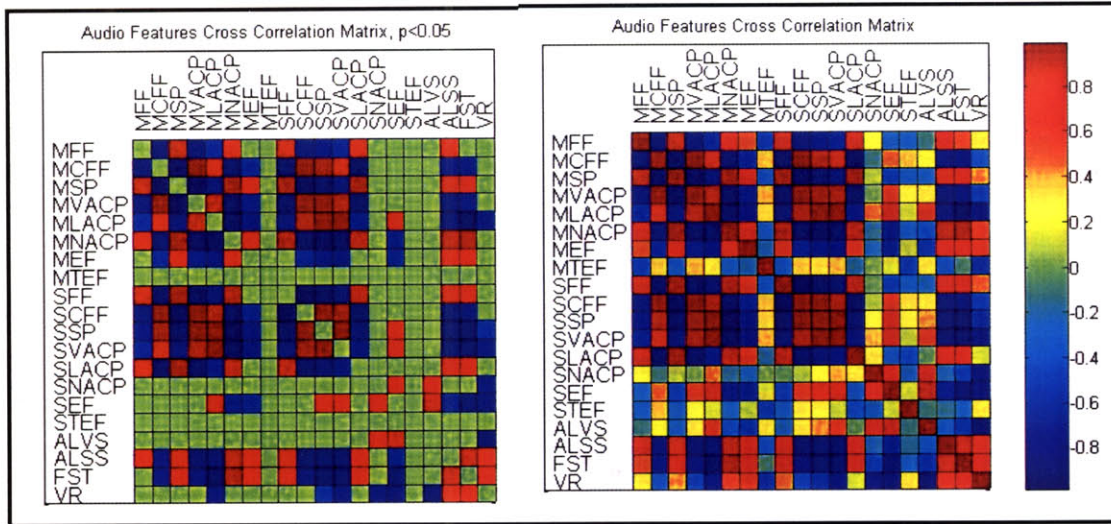
Emotion Rating	Multivariate Regression			Emotion Rating	Multivariate Regression		
	R <sup>2</sup>	$\beta_1$	$\beta_2$		R <sup>2</sup>	$\beta_1$	$\beta_2$
Depressed	0.80	23.1	49.3	Happy	0.85	-25.6	-41.9
Stressed	0.23	13.0	8.1	Frustrated	0.09	11.5	-10.7
Mentally Alert	0.37	6.2	-11.1	Sleepy	0.11	-13.3	3.2
Anxious	0.21	-13.9	2.0	Angry	0.23	12.5	-13.4
Guilty	0.62	34.2	34.7	Distracted	0.01	3.6	-2.6
Self-worth	0.89	-26.4	-51.0	Sad	0.73	40.9	38.8
Emotional	0.41	21.9	21.7	Aroused	0.47	-29.6	-23.2
Optimistic	0.83	-27.0	-50.0	Afraid	0.33	-14.4	-12.2
Energy Level	0.85	-20.5	-46.9	Disgusted	0.08	6.2	-1.6
Hopelessness	0.75	18.7	43.3	Bored	0.30	0.1	-0.2
Excited	0.80	-41.6	-32.8	Embarrassed	0.29	11.0	-1.9
Apathetic	0.23	-12.1	-10.8	Calm	0.32	2.5	-18.3
Sociable	0.50	-15.2	-29.4	Tired	0.27	-4.2	19.1
<b>Negative Emotion</b>	<b>0.79</b>	<b>1.7</b>	<b>2.5</b>	<b>Positive Emotion</b>	<b>0.85</b>	<b>-2.5</b>	<b>-4.1</b>
<b>Composite Emotion</b>	<b>0.84</b>	<b>-4.2</b>	<b>-6.6</b>				

**Table 8: Chart showing multilinear regression using the top two features (std. dev. of the value of the largest autocorrelation peak in the audio, and max. value of the accelerometer), across the 26 dimensions of self-reported emotions as well as for the depression emotion indices for Subject 6. The R<sup>2</sup> of all the emotion indices are between 0.79-0.85.**

## Feature Dimensionality Reduction Analysis

Many of the potential input features used in the study have many cross-correlations with each other. The following shows the correlations between the physiology and voice features:

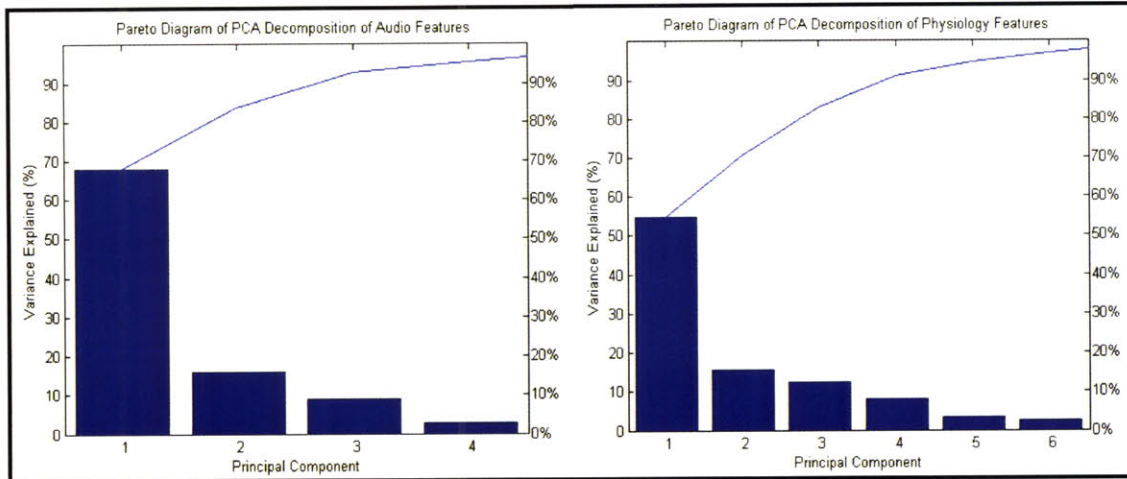




**Figure 38: Cross-correlations between the voice features. The features are: Average length of voiced segment (ALVS), Average length of speaking segment (ALSS), Fraction of speaking time (FST), Voicing rate (VR), Mean/standard deviation of the fundamental formant frequency (MFF/SFF), Mean/standard deviation of the confidence in the fundamental formant frequency (MCFF/SCFF), Mean/standard deviation of the spectral entropy (MSE/SSE), Mean/standard deviation of the value of largest autocorrelation peak (MVACP/SVACP), Mean/standard deviation of the location of largest autocorrelation peak (MLACP/SLACP), Mean/standard deviation of the number of autocorrelation peaks (MNACP/SNACP), Mean/standard deviation of the energy in frame (MEF/SEF), Mean/standard deviation of the time derivative of energy in frame (MTEF/STEF), Conversation Level (CL), Stress Level (SL).**

For the voice features, note that there are a variety of other candidate features that could have been chosen by the stepwise selection: the MFF, MCFF, MSP, MVACP, MNACP, SFF, SCFF, and SSP features all have a very high correlation to SLACP in the range of 0.82-0.95 (absolute value). Likewise, for the physiology features, many of the features are highly correlated to each other. In particular, the various accelerometry features (std. deviation, sum of accumulated differences, entropy, energy, maximum) all had very high correlations ( $R = 0.85-0.99$  range) to each other.

Given these cross-correlations, it could be desirable to use dimension reduction techniques to replace the large set of features with a smaller set of new independent features that hopefully have as much descriptive power as the original feature space. In order to do this, we do a principal component decomposition analysis on the physiology and audio feature spaces to see if we can find an appropriate set of features in each category of features. The following shows the Pareto diagrams for the resulting PCA decompositions:



**Figure 39: Pareto diagrams of PCA decomposition of audio and physiology features. We see that the first principal component in both cases explains the majority of the variance of the data, and thus is a good candidate for a potential feature to use in predicting the emotion ratings.**

We see from the diagrams that the first three audio principal components roughly explain 68%, 16%, and 9% of the variance of the audio features, and the physiology principal components explain roughly 55%, 15.8, 12.3% of the physiology features. In both cases, the first principal component explains a majority of the variance in each feature space, and thus is a good candidate for a potential feature to use in predicting emotion ratings.

Using the first audio principal component in a regression against the composite emotion index, we find a  $R^2$  of 0.40,  $p = 0.045$ . Likewise, using the first physiology principal component results in a regression with an  $R^2$  of 0.38,  $p = 0.037$ .

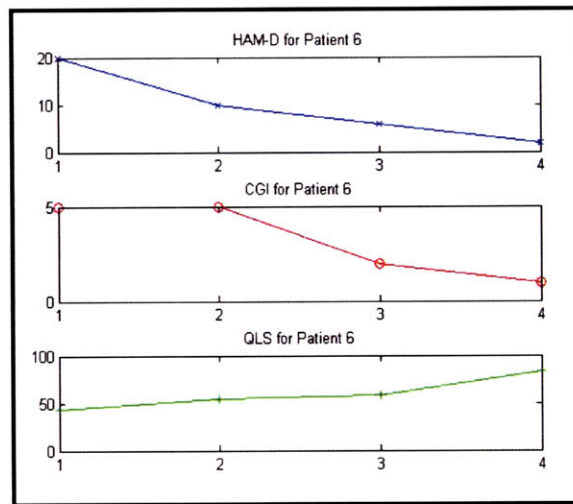
We take these two top components from the audio and physiology decompositions as input features into the multi-linear regression on the composite emotion index. This results in a regression with an  $R^2$  of 0.59, with betas of  $\beta_a = -1.78$ ,  $\beta_p = 0.78$ . These results are not quite as good as just using the top two raw features chosen by the stepwise linear regression (this turns out to be true for all four of the subjects who have completed the study protocol so far).



## Clinical Outcome Trending

Now that we have discovered features that highly correlate to subject's self-reported emotion ratings, we can now see if these features are also highly correlated to the clinical outcome scales that were obtained at week intervals.

The clinical outcome scales (HAM-D/CGI/QLS scales) are administered by a psychiatric physician at the start and end of a subject's stay at MGH, as well as at the end of every week of stay. Thus we have relatively course-grained snapshots of the clinical mental health of each subject over the duration of each treatment stay. The clinical outcomes for Subject 6 are summarized below:



**Figure 40: Weekly clinical outcomes over 3-week period for Subject 6 (for CGI, CGIsev was used). We see that the ECT therapy has successfully treated the subject, reducing a severe clinical depression to something that is below a mild depression as rated by the HAM-D. Correspondingly, the CGI severity score decreases and the QLS increases over the same period.**

The HAM-D is the standard measure used in the psychiatric field to diagnose depression. A rough guide is a HAM-D score of greater than 20 signifies severe depression, a score greater than 15 indicates moderate depression, and a score greater than 8 indicates mild depression. For Subject 6, we see that she starts out with a severe clinically-diagnosed depression, falls to a mild depression after the first week of treatment, and is well on her way to recovery with a HAM-D score of 6 by the end of week 2.

The CGI is another clinical measure for assessing the depression level of a patient, and ranges between 3 (no depression) to 21 (very high depression). For Subject 6, the composite CGI scale was not obtained for the start so we use the severity measure subcomponent (which ranges from between 1 to 5) as proxy.

The QLS is a self-reported questionnaire asking about a patient’s quality of life across a variety of dimensions of daily life (ranging from physical, social, work, household, classes, and leisure activities). Because some activities are not applicable in the in-ward setting at the hospital, instead of using the strict value of the QLS measure, we calculate a scaled percentage metric based on the ratio of the raw score to the maximum total score based on applicable questions. We see that the Subject 6’s QLS scores under 50% the first week, improves slightly in weeks 2 and 3, and then jumps up at the end of her treatment.

An interesting question to ask is how the subject’s self reported data and derived composite emotion indices correlates to these clinical measures. These correlations are shown in the correlation matrix below:

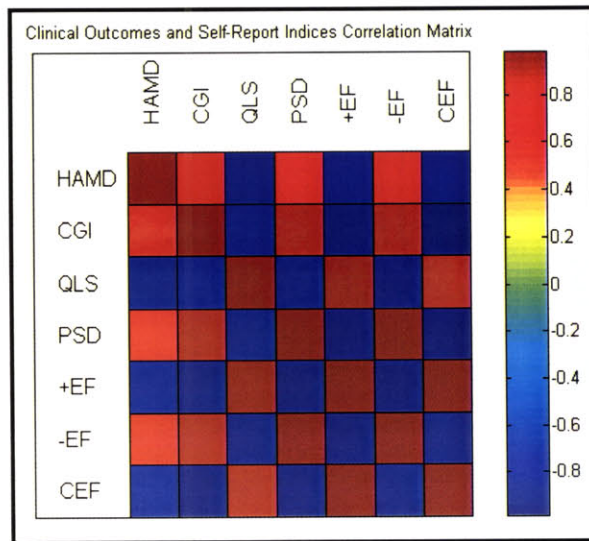


Figure 41: Correlation matrix of clinical outcomes and self-reported emotion indices (PSD = patient self-reported depression, -EF = negative emotion index, +EF = positive emotion index, CEF = composite emotion index). All correlations are in the range between 0.82-0.99.

We see that there are very strong positive correlations between the HAM-D, CGI, and negative emotion factor, as expected. Likewise, there are strong correlations between the QLS, the positive emotion factor, and the composite emotion factor (which was defined to be a positive emotion factor). These correlations indicate that the features that are relevant for the composite emotion indices are also very relevant for predicting the clinical outcomes as well.

Using the top two features picked for the composite emotion index, we get the following regression results when run on the clinical outcomes.

Clinical Outcome	Multivariate Regression		
	R <sup>2</sup>	$\beta_1$	$\beta_2$
HAM-D	0.69	-0.02	2.48
CGI	0.94	-6.7	52.4
QLS	0.49	0.45	-2.82

**Table 9: Clinical outcome regression results for Subject 6 using top two features chosen in stepwise regression analysis for the composite emotion index, along with their betas.**

While these regression results are not quite as good as with the self-reported emotion indices, they are still fairly accurate.

### ECT vs. non-ECT days

One important aspect of the analysis thus far ignored was the fact that ECT treatment can result in changes in emotions and physiology. There is not a complete theory of why ECT treatment works, despite the fact that ECT is one of the most effective treatments known (80% of subjects who do not respond to medication and other treatments become significantly improved following ECT). ECT works by running a current through the head, essentially shocking the brain and inducing a controlled seizure. While there are no serious long-term effects, immediately after treatment ECT can cause somatic side effects such as nausea, tiredness, muscle aches and other physical discomfort as well as cognitive side effects such as headaches, disorientation/confusion, and short-term memory loss.

For the subjects in the protocol, ECT treatment is normally initiated three times a week, every Monday/Wednesday/Friday in the early morning during the length of stay in the ward. Given the potential side effects of the treatment, it is very plausible that ECT can cause the mood of subjects to change temporarily. The effects of ECT vary between subjects, we find the corresponding short-term swings in emotions to vary as well. Some subjects, such as Subject 6, did not exhibit negative symptoms related to the ECT treatment. Other, such as Subject 3, reacted to the ECT treatment which resulted in temporary shifts in emotion, typically registered during the days of treatments right after an ECT session. These changes in emotion can be, at the least, partially explained by the physical discomfort reported by the patient.

Because the LiveNet systems were worn 24-hours a day, the monitoring systems were even present during the actual electroshock therapy sessions. This offers the unique ability for us to monitor the physiological effects of ECT therapy before, after, and even during the treatment process itself. However, despite the potential shifts in physiology that occur during the ECT therapy itself and the short term mood swings that can result due to the side effects, we could find no strong relation correlating the aggregate physiology between ECT days and non-ECT days. Correlations comparing the physiology for the subjects were found between 0.09-0.15, or slightly better than chance. Non-supervised clustering analysis was also unable to partition the observational days based on the voice or physiology features alone (in contrast to using these features to cluster the depression state of the individual), indicating that outside of short-term changes, there were no aggregated physiology shifts that could be detected at the granularity of a day. This perhaps is a good thing in terms of the subject, indicating that there are no long-term physiology changes as a result of ECT treatment and any changes in physiology can be accounted for by the improvement in depression state.

In fact, existing research seems to confirm this finding; studies investigating the side effects of ECT could find no long-term cognitive effects following ECT, and any subjective somatic side effects during the ECT course did not change from pre-ECT measurements, suggesting that somatic side effects are related to depressive state rather than being induced by ECT [Devanand et al. 1995]. Another study corroborates this finding, reporting that the subjective side effects experienced by patients is related to depression level, and improvement in depression correlated with corresponding reductions in side effect burden [Brodsky et al. 2001].

## Correlating Behavioral Features

It is very clear from the literature described in Section 7.2 that depressed patients readily exhibit behavioral changes as well as changes in socialization patterns. In addition to the raw physiology and audio features that were recorded there are some obvious behavioral features that are readily derived from the data collected in the study. Using the LiveNet system, we can quantify many behavioral patterns and use these as features to correlate with the emotional and outcome measures of interest. The behavioral measures we derive include the activity fraction (fraction of the day spent doing physical activity), sedentary fraction (fraction of the day spent in a sedentary state, i.e. not moving much but not sleeping), sleep fraction (the fraction of the day sleeping), restlessness fraction (average rate of flipping while sleeping), and in-room fraction (the fraction of the day spent in the room).

The results for the behavioral features, as correlated to the Composite Emotion Index is shown in the Table 10 below.

Behavioral Top Features	Regression Statistics	
	R <sup>2</sup>	p
1) Sedentary Fraction	0.64	0.012
2) In-Room Fraction	0.58	0.023
3) Activity Fraction	0.22	n.s.
4) Sleep Fraction	0.15	n.s.
5) Restlessness Fraction	0.09	n.s.

**Table 10: Behavioral features correlations with Composite Emotion Index for Subject 6. Sedentary behavior was the best behavioral predictor for the subject, followed by the in-bedroom measure as determined by the location beacons.**

Thus we see that behavioral features can correlate well with outcomes. The sedentary (the amount of time spent in a sedentary state) and in-room location features perform the best, with correlations that are almost as good as the top voice and physiology features. The sedentary

fraction and in-room fraction features are highly correlated, but this is probably due to the fact that within the room in the ward setting, there isn't anything to do but really lie in bed to watch television. We would expect that in a more general ambulatory setting, these two features would potentially be different, with sedentary behavior more correlated with lack of energy and enthusiasm, while staying within your room may have a socialization component.

Although it is expected that activity fraction behavioral measure and the motion features aggregating motion characteristics above are correlated, apparently the activity feature is not as correlated to the emotion factor. The activity features doesn't perform as well in this setting, but that is probably due to the fact that patients are in a ward with very limited options besides moving around from area to area, so there is very little activity overall.

Sleeping and restlessness features are correlated the least; in this setting, it appears that the amount of sleep and restlessness is not correlated to depression state (i.e., people sleep about the same regardless). A larger study population size is required to determine if changes in sleep and or restlessness can be correlated to depression level.

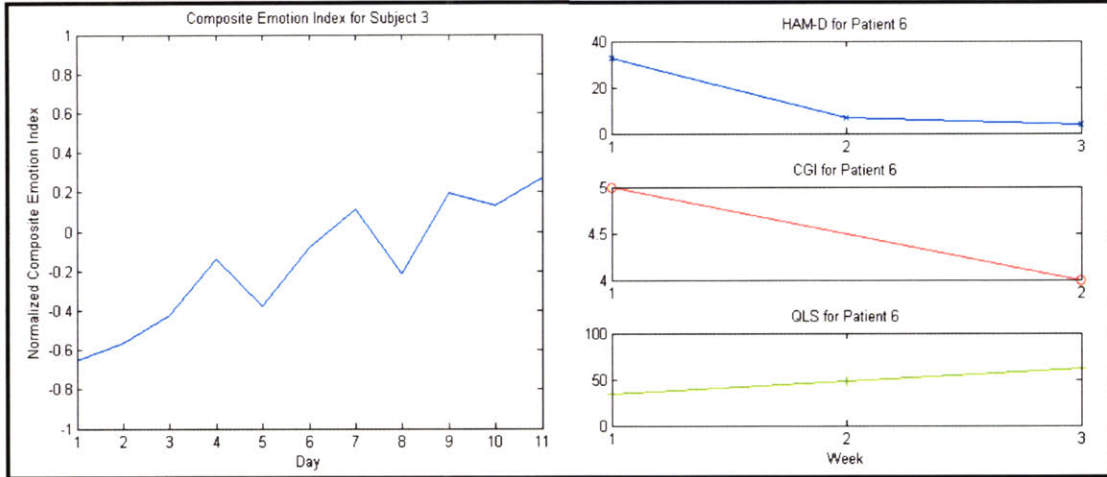
In general, we find that while the behavioral features are quite correlated to outcomes, they do not outperform the physiology and voice features when using these features for regression.

## **Longitudinal Subject Results**

In the above analysis, we have shown that it is possible to use aggregated physiology and behavioral features to predict depression state, based on the data for Subject 6. We now use the same methodology on the data of the other subjects run so far in the study.

### **Analysis for Subject 3**

Subject 3, a woman in her mid thirties, was the second subject run through the protocol to completion, but the first that resulted in a full set of data. Subject 3 was only on the ward for a week and a half, but treatment was highly effective for her. It was found that Subject 3 exhibited relatively large changes in mood on a number of her ECT treatment days, which she attributed to physical discomfort following treatment. She was the only patient run through the protocol thus far who exhibited these large fluctuations in mood.



**Figure 42: Subjective composite emotion index and HAM-D, CGI, and QLS clinical outcomes for Subject 3. Subject 3 was only on the ward for 11 days, and showed steady improvement from the ECT therapy as shown in both the subjective and clinical outcomes. She was released to an outpatient treatment program at the end of the second week.**

Here we see that there is a characteristic dip in the Composite Emotion Index for Subject 3 that occurs periodically on ECT days. There is a general improvement trend for Subject 3, as noted by the Composite Emotion Index as well as the clinical outcomes reported (note that the Composite Emotion Index is at the granularity of a day, and the clinical outcome measures are weekly).

For Subject 3, we first use stepwise regression analysis to find the top features correlated to the composite emotion index:

Subject 3 Top Features	Regression Statistics	
	R <sup>2</sup>	p
1) SVACP	0.52	0.013
2) Std. Dev. Acceleration	0.41	0.024
3) SLACP	0.38	0.028

**Table 11: Top features selected for multivariate regression against composite emotion index (chosen with a p<0.05 criterion) for Subject 3. Std. dev. of the value of the largest autocorrelation peak, std. dev. in motion, std. dev. in location of the autocorrelation peak were the top three performing features.**

The best features chosen in the stepwise regression involved the standard deviation of the value of the autocorrelation peak and the standard deviation in the location of the autocorrelation peak (again, pitch variation measures) as well as the standard deviation in the accelerometer (which measures motion variation). The behavioral features for Subject 3 were not very discriminating at all, since the subject was essentially in a sedentary state for the duration of her stay except when forced leave the bed by study staff. The activity and sleep features did not correlate well with outcomes either.

Using the top feature from each class, we generate the following regression results for the composite depression indices and clinical outcomes:

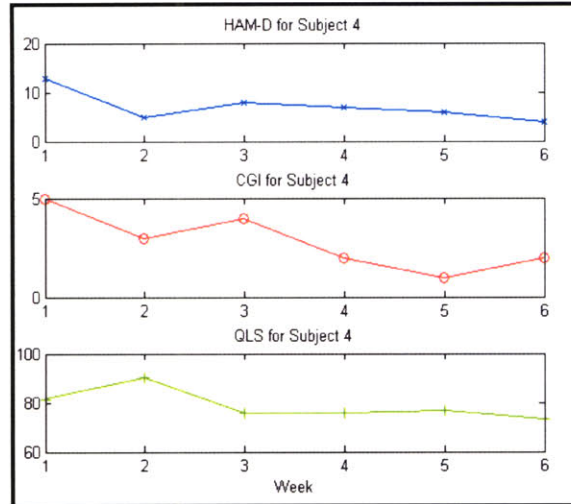
Subject 3 Clinical Outcome	Regression Statistics		
	R <sup>2</sup>	$\beta_1$	$\beta_2$
Composite Emotion Index	0.70	2.64	-1.40
HAM-D	0.63	-0.22	1.68
CGI	0.87	-4.3	37.7
QLS	0.55	0.97	-6.48

**Table 12: Subjective and clinical outcome regression results for Subject 3 along with their betas. All correlations for a two feature model have an R<sup>2</sup> between 0.55-0.87.**

#### **Analysis for Subject 4**

Subject 4, a man in the mid fifties, was the second subject run through the protocol to completion. Subject 4 was very unique in that he stayed on the ward for treatment for over 8 weeks, which is an unusually long length of time (most patients leave after a few weeks of treatment). He initially responded to treatment in the first 2 weeks, but his condition as reported by his subjective and clinical outcomes did not improve after week 3. A psychiatrist mentioned that perhaps he was not actually getting more depressed but was misinterpreting anxiety as depression.





**Figure 43: HAM-D, CGI, and QLS clinical outcomes for Subject 4. Subject 4 stayed on the ward for over 8 weeks, and initially showed improvement from ECT therapy treatment through the second week, but progress then stalled until after the fifth week**

We see from the clinical outcomes that after an initial improvement from depression, subject 4 actually becomes more slightly more depressed in Week 3 and then stabilizes without much improvement thereafter.

Unfortunately, some of the timestamps for emotion rating survey data for Subject 4 was corrupted, so we were not able to accurately reconstruct the aggregated daily emotion ratings. As such, only the clinical outcomes regressions are conducted. Performing a stepwise regression on the HAM-D, we obtain the following features:

Subject 4 Top Features	Regression Statistics	
	R <sup>2</sup>	p
1) SFF	0.56	0.011
2) Motion Energy	0.47	0.023
3) Motion Entropy	0.42	0.031

**Table 13: Top features selected for multivariate regression against composite emotion index (chosen with a p<0.05 criterion) for Subject 4. Std. dev. of fundamental formant, motion energy, and motion entropy were the top three performing features.**

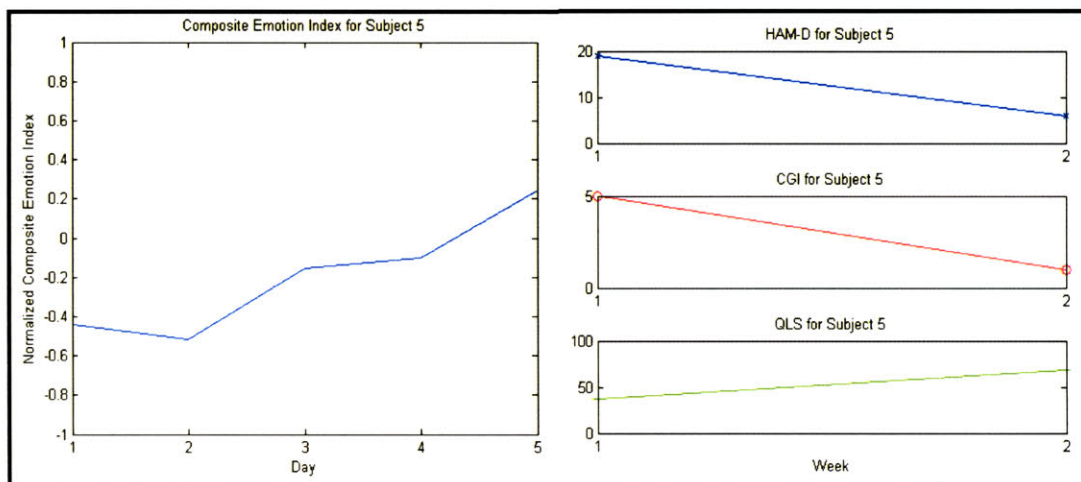
Again, we see the reoccurrence of pitch variation (standard deviation of fundamental formant frequency) and measures of the entropy and energy in the motion as the top features. Using the two top features, we generate the following regression results for the composite depression indices and clinical outcomes:

Subject 4 Clinical Outcome	Regression Statistics		
	R <sup>2</sup>	β <sub>1</sub>	β <sub>2</sub>
HAM-D	0.60	-1.5	5.4
CGI	0.89	-12.4	49.3
QLS	0.47	2.7	-10.5

**Table 14: Subjective and clinical outcome regression results for Subject 4. All correlations for a two feature model have an R<sup>2</sup> between 0.47-0.89.**

### Analysis for Subject 5

Subject 5, a male in his around forty, responded very well to treatment, and was out of the ward within a week and a half, as shown in the figure below:



**Figure 44: Subjective composite emotion index and HAM-D, CGI, and QLS clinical outcomes for Subject 5. Subject 5 stayed on the ward for only one and a half weeks, but showed improvement to ECT therapy treatment throughout his stay on the ward.**

For subject 5, we first use stepwise regression analysis to find the top features correlated to the Composite Emotion Index:

Subject 5 Top Features	Regression Statistics	
	R <sup>2</sup>	p
1) SLACP	0.79	0.003
2) Std. Dev. Accel.	0.86	0.0028
3) Max. Num. Peaks Accel.	0.63	0.015

**Table 15: Top features selected for multivariate regression against composite emotion index (chosen with a  $p < 0.05$  criterion) for Subject 5. Std. dev. of the value of the largest autocorrelation peak, std. dev. in motion, maximum number of peaks in motion were the top three performing features.**

Using the top feature from each class, we generate the following regression results for the composite depression indices and clinical outcomes:

Subject 5 Clinical Outcome	Regression Statistics		
	R <sup>2</sup>	$\beta_1$	$\beta_2$
Composite Emotion Index	0.72	1.8	-3.9
HAM-D	0.84	-0.7	5.3
CGI	0.95	-10.3	72.4
QLS	0.73	3.6	-8.3

**Table 16: Subjective and clinical outcome regression results for Subject 5. All correlations for a two feature model have an R<sup>2</sup> between 0.72-0.95.**

As many of the audio and physiology features are highly correlated, it is not surprising that individual features chosen in the stepwise regression of each subject are not the identical to each other. In addition, idiosyncrasies between subjects are to be expected. However, we would like to try to generalize the results of the study to see if any universal features exist across the subjects that can be used to accurately trend depression.

## Toward Creating a Universal Depression Meter

The above analysis shows that we can accurately predict an individual's depression state based on both a combination of voice, physiological, and behavioral features. While it is always possible to create a personalized depression model based on an individual's physiology and behavior, an interesting question to ask is if these same features hold across patient populations. While an appropriately powered longitudinal study is not possible at the time of the writing of this thesis (with only 6 subjects run through the protocol), we can see if there are universal features that can be used for clinical depression state classification.

Taking the top performing features of each individual subject (discussed in the longitudinal results section above) and systematically running the regressions with these features on the other subjects, we can get a first pass sense as to what the best performing feature set is across all of the subjects. In general, we find that the two general classes of features that perform the best were the ones that measure the pitch variation in the voice (SVACP, SLACP, MFF) as well as motion energy and variation features. The best performing combination across subjects takes one voice feature and one motion feature (std. dev. location of the auto correlation peak and std. dev. of acceleration). The results are summarized in the table below:

Subject #	Regression Statistics ( $R^2$ )		
	HAM-D	CGI	QLS
Subject 3	0.57	0.77	0.46
Subject 4	0.56	0.81	0.43
Subject 5	0.84	0.95	0.73
Subject 6	0.62	0.83	0.51

**Table 17: Top feature sets performance across subjects using the std. dev. of the location of the autocorrelation peak and the std. dev. in the raw motion features. HAM-D outcomes can be correlated with an  $R^2$  between 0.57-0.84, CGI with an  $R^2$  between 0.77-0.95, and QLS with an  $R^2$  between 0.43-0.73.**

It is interesting to note that the voice pitch variation and motion features are the dominant performing features across all subjects. Features such as skin conductance, heat flux, were not nearly as discriminating, and do not make it into the top feature list of any of the subjects. From the data, it appears that we can consistently trend the CGI the best, and QLS is the least accurate based on the features. This makes sense as the features were chosen to optimize the composite emotion index (which correlates the most to the HAM-D and CGI, which are clinical depression measures).

## **7.6 MGH Study Discussion**

This MGH study serves to establish that the LiveNet system is capable of engaging in significant long-term ambulatory clinical studies. The subjects were able to comfortably wear the monitoring system continuously daily for long periods of time, in one case for over eight weeks. This allows the ability to gain compelling physiological and behavioral data in potentially completely naturalistic settings. None of the patients complained about discomfort while wearing the variety of sensing and monitoring gear, and all reported that they grew accustomed to wearing the system very quickly. The reliability and robustness of the LiveNet infrastructure had been improved and demonstrated under sustained use. In addition, the systems proved to be easy to use, even such that patients could put on and use the system without any other guidance or supervision.

The implications and clinical significance of the research are broad. The study is the first to our knowledge to combine continuous behavioral and physiologic measures in a wearable device in a study of clinical depression. The development and refinement of a methodology that objectively and accurately monitors treatment response in major depression has implications for diagnosing clinical depression, measuring response to treatment, and promoting relapse prevention.

First and foremost, this research sheds more light on the physiology of depression and improves our understanding of how the role of the autonomic nervous system affects physiology and behavior. This has major implications toward creating an objective standard for depression that has thus far eluded clinical diagnosis. We show that it is possible to create an individualized objective metric for depression based on simple physiological measures. Furthermore, we have evidence that there are universal physiological features that can potentially be used to create a universal objective metric for depression. While the results for the subjects given thus far are

promising, it remains to be seen whether these preliminary results will hold up in the context of the continuing large-scale longitudinal study. Ultimately, the research will give more insight to researchers so that they can develop more sophisticated physiologic models that describe the mechanism of healing, which will improve the efficiency and efficacy of treatment. We can even begin to delve into differentiating between the subtypes of depression and endophenotyping for genetic studies of depression can be studied in ambulatory settings.

We have been able to collect a novel dataset of physiology, behavior, and emotional data over long-term periods. From the standpoint of just the emotion data, we have explored how a wide spectrum of emotions for depressed patients vary over time and how they co-relate. We demonstrate that an eigendecomposition of the emotion rating data can provide insight on how to classify various emotions. The interpretation of this decomposition is, for the most part, consistent with known models of emotion theory. Much as eigenface decomposition provided a powerful way of deconstructing faces for facial recognition applications [Turk & Pentland 1991], this methodology can allow us to reconstruct more complicated emotions based on more fundamental underlying emotions. This may be beneficial to compare and perhaps even identify underlying neurobiologic mechanisms controlling certain emotions.

Another important finding is the fact that we only need to look at the aggregate physiology and behavior to be able to gauge the general level of depression of a patient. In fact, the results point to the fact that longer timescales actually improve the correlation with the clinical outcomes as well as subjective emotion ratings. We find that the aggregated physiology at the granularity of a day does a good job of trending general depression levels. Thus, it is not necessary to run sophisticated models or look at detailed time dependencies in the data to accurately predict depression levels from the physiology and behavioral data. In fact, standard linear classifiers based on simple measures such as aggregate motion energy or variation in voice pitch have been shown to perform remarkably well in tracking depression state throughout the entire treatment process. Furthermore, these features can easily be calculated in real-time locally on embedded hardware such as mobile phones and PDAs, reducing the system requirements for a practical system for monitoring depression.

Further, we have established that we can quantitatively characterize long-term behavior. Although the behavioral features that were derived did not end up performing quite as well as the

top physiology and voice features, a number of them are significantly correlated to the depression outcomes. In general, many psychological and physiological disorders have symptoms that manifest as behavioral changes. Unfortunately, without a way to precisely quantify these behaviors, physicians can only qualitatively describe these behaviors. With continuous monitoring, we now have a way to develop objective behavioral measures which can be used to help diagnose and/or classify disorders. The behavioral features used in this study, such as activity, sleep, sedentary measures, are generally quite applicable in a wide assortment of contexts. As these measures can all be derived in a completely non-invasive manner (through the use of a single body-worn accelerometer), they may be attractive in developing depression classifiers in situations where other sensors such as microphones might be more cumbersome.

Once a reliable index of physiologic and behavioral metrics for depression has been established, we also have an automated methodology to be able to track patient response to treatment over time. As a tool for monitoring clinical change, these measures are invariant to the type of treatment, be it psychotherapy, pharmacotherapy, ECT, or combination thereof. The degree of change in the developed measures can be directly related to the degree of response to the treatment. Thus, by monitoring pre and post treatment, therapeutic changes can be measured and as each subject acts as their own control for such purposes, objectivity is assured.

Furthermore, this monitoring can be done outside of the inpatient setting in completely natural settings, i.e. at the patient's home, without the need for additional intervention or support. The preliminary results of this study demonstrate the tantalizing possibility of being able to accurately monitor and track depression using a mobile phone with an integrated accelerometer. While a full PDA-based LiveNet system with sensor network was used in the study to collect the variety of physiology and behavior measures, our results show that we can accurately correlate depression to simple aggregated motion and voice features. By using a mobile phone, we thus have a platform that can be used to track depression while being completely transparent to the user. For individuals who are prone to relapses and severe cases of clinical depression, this ability to continuously monitor patients when they leave the ward is critical in being able to expand healthcare into the homecare domain where existing treatment methodologies have not yet demonstrated substantial effectiveness as compared to in-ward settings.

As we improve our understanding of how the autonomic nervous system affects long-term physiology and behavior, we can begin to look at other potential mental disorders. The methodology that we have developed also has the potential to help in the pre-clinical assessment, early diagnosis, and treatment prediction of other severe psychopathologies that have likely physiologic and behavioral correlates. These conditions include severe psychotic, mood, anxiety, communication, and personality disorders, and potentially even pervasive developmental disorders in children. For example, an easy target for assessment would be bipolar disorder, where a patient exhibits extreme swings in mood related to manic and depressive episodes. These episodes have corresponding swings in behavior which can easily be identified and quantified based on movement and speech patterns.



## Chapter 8 Long-Term Behavioral Modeling

As is demonstrated in the MGH Depression Study, the LiveNet platform lends itself naturally to be the ability to do a wide variety of long-term clinical healthcare monitoring applications. However, this type of technology is also relevant outside the traditional healthcare and clinical contexts. It is very important from the proactive healthcare point of view that monitoring systems also be able to determine trends in physiological/contextual state over time to provide not only immediate diagnostic power but also prognostic insight into the future condition or behavior of an individual.

The continuous behavioral and contextual data such as of the type collected in the MGH Depression Study can be used to identify even longer-term trends in behavior that may span months or even years. By looking at the long-term patterns of a person's high-level activity behavior, we can extract the texture of their everyday life, which we call a person's 'lifescape'. We demonstrate that it is possible to develop relatively accurate time dependent models of a person's lifescape using HMM modeling techniques described briefly in Section 4.10). Thus, we can use the texture of physiological and contextual information from a person's routine high-level behavior over time to profile their normal living style and then detect trends in changing behavior based on these models.

In order to study and model normal human behavior, we must first collect data about the long-term daily routines and activities of individuals. For the Lifescape Study, we have collected the long-term everyday behavioral data of a number of subjects, including the author. The rest of chapter describes the long-term trending and behavior modeling work that has been attempted based on this dataset.

### 8.1 Lifescape Study Background

The LiveNet platform also lends itself naturally to be able to do a wide variety of long-term healthcare monitoring applications for physiological and behavioral trends that vary slowly with time by using the currently available physiological sensors. Most existing health monitoring

systems are purely reactive (e.g. sounding an alarm after a person has a heart attack or falls down), and focus on classifying and determining in real-time when certain events have occurred.

While this type of application is very useful and potentially-life saving, these systems typically do not have any sense of the history of an individual and can only react to instantaneous events. By combining long-term monitoring with context classification, it is possible to develop more proactive systems and personalized data that can be used to catch problems before they manifest themselves (e.g. instead of reacting to a heart attack, one can predict beforehand that a heart attack is imminent). However, a proactive system requires more resources, as it must have context-aware and inference capabilities to be able to determine what the right information is to be directed at the right people, to the right places, at the right times, and for the right reasons. While challenging, small advances have been made in this regard.

With systems that can potentially store your long-term activity and health information, it opens the door to long-term physiological trending analysis as well as being able to track activity to detect potentially important deviations from normal healthy behavior. Companies such as Living Independently have attempted to develop systems that can keep track of elderly people living at home, and alert family or healthcare professionals if behavior suddenly changes. The Living Independently system works by using infrared motion sensors that can track people moving about the house. Sudden changes in behavioral patterns (such as not leaving bed by 2 pm when one normally wakes up at 8 am) could trigger an alarm. A similar idea using low-cost accelerometer-based activation sensors, which can be spread all over a living environment (under the couches, in drawers, cabinets, doors, etc.) has been implemented by the MIT Media Laboratory [Tapia et al. 2004].

Specifically in the realm of elder care monitoring, which is one of the primary motivations for long-term behavior modeling, being able to determine activities of daily living is key to assessing the status of the elderly, as it has been shown to be a good predictor of nursing home acceptances and mortality [Wiener et al. 1990]. There is a large research effort to identify new technologies that are able to automatically track and classify activities of daily living [Lynch et al. 2003]. Most of these techniques have focused on specifically identifying activities that are important for the elderly to be able to function and live independently (such as taking medication, cooking, etc.) and focuses on instrumenting the environment with a wide variety of sensors, oftentimes with

very expensive equipment (including video, audio, activity switches, etc), such as the Smart Home at The University of Rochester's Center for Future Health or the TIAX PlaceLab.

In contrast, we attempt to conduct health trending and high-level human behavioral modeling using only minimal sensing infrastructure in the form of a wearable system. Given the advances in low-power processing, battery densities, wireless sensing capabilities just over the horizon, wearable technology will become an ever more compelling platform for healthcare monitoring applications. In fact, we are already beginning to see this convergence; in early 2005, Samsung announced the world's first mobile phone with a built-in accelerometer which can be used as a gesture interface. Nokia and other phone manufacturers are following suit, resulting in mobile platforms which already have all the necessary components to enable behavioral analysis. As technology shrinks and becomes less and less unwieldy and invasive, we are moving closer towards a world where it may become commonplace for people to wear long-term health monitoring systems even when they are not sick.

Based on the fact that people tend to be creatures of habit, it is possible to monitor repetitive behavior over long periods of time to train models, which can then be used to predict future behavior. Studies in long-term behavioral context classification have been attempted in the past. In Brian Clarkson's "I Sensed" Project, one hundred days of video/audio data from his life were collected and analyzed for long-term patterns of behavior. From the research, it was shown that people tend to be very repetitive in terms of their daily behavior. Clarkson's research demonstrated that long-term behavioral trending and analysis could be performed using only basic 32x32 pixel video features consistent with the granularity of insect vision [Clarkson 2002].

Similarly, we wish to demonstrate that we can use non-invasive sensing to model the behavior of people going about their daily lives. These technologies allow us to move to a more proactive healthcare paradigm, in which computer-assisted technology can help to track personal health trends and monitor and correct non-healthy behaviors before they become chronic (in addition to the standard critical health monitoring applications that we focus on today).

As shown in previous chapters, movement has proven to be a highly effective way capturing the activity patterns and behavior of individuals. Likewise, thermoregulation patterns of normal temperature and heat flux fluctuations that people exhibit through their circadian cycles also can

be used as non-invasive and characteristic markers for high-level behavior trending. We show that we can use these types of sensing data to develop time-dependent models of human behavior over long-term timescales.

## 8.2 Lifescope Experimental Protocol

For this study, we wish to track the long-term behavior of subjects going about their daily lives. In order to do this, we use a long-term armband monitor that is capable of tracking the motion, skin conductance, heat flux, and temperature. This armband sensor was worn by every subject 24-hours a day, 7 days a week for the duration of the data collection period, which ran from May, 2004 until August, 2005. The only time the armband is taken off by a subject is during showers or other activities involving water contact.

Because of the memory limitations of the armband, this physiological/behavioral data was sampled at a relatively low sample rate of 30 samples/min. If the armband sensor data could not be collected on a daily basis, such as weekends and longer trips away from the collection site, the armband sampling rate was adjusted to the maximum rate that still allowed the continuous recording over the period (this sampling rate was set at 15 samples/min on weekends, and never lower than 1 sample/min).

### Behavior, Location, and Interaction Annotation Log

For the study, each subject kept a daily diary of the salient events, activities, and contexts as they went about their everyday lives. As such, the subject diaries recorded detailed behavior annotation and timestamp information regarding the daily life of the subjects. These diaries included location information (e.g., ‘in office’, ‘at home’, ‘at Anna’s Taqueria’, ‘at the Beacon Hill Pub’), detailed activity and context information (e.g., ‘working’, ‘watching t.v.’, ‘sleeping’, ‘socializing at bar’), as well as interaction information with other individuals (e.g. ‘with Jon and Mike’, ‘meeting with Sandy’). Every context is accompanied with timing information that denotes the start and end of a particular activity/context, with a resolution of approximately 5-10 minutes. In a preponderance of cases, the interesting changes in context can be accurately time-stamped by simply observing the raw physiological data stream by inspection and noting basic changes in accelerometry and heat flux spikes. Manual transcription of the raw behavioral and

contextual information contained in the diaries has been initiated to assign unique and consistent codes to each type of activity/context as well as to synchronize the annotations to the behavioral data, to allow for future analysis.

## **Cell Tower Location IDs and Bluetooth Interaction Logs**

In order to corroborate diary entries for location for our baseline analysis, we can use cell tower ID information. In addition to the armband monitor, each subject carried a Bluetooth-enabled mobile phone used in a social-networking study called Serendipity [Eagle 2005]. Using logging software developed from Serendipity, we are able to continuously log all the IDs of the cell tower regions that a subject traversed through. These cell tower IDs have approximately the granularity of region-sized areas in Metropolitan Boston, which can be equated to wider-area geographical location information of the subject. For the subjects involved in the study, this allows us to give unique cell-tower IDs to differentiate locations such as home and work.

The mobile phone also recorded all IDs of any Bluetooth devices that were nearby. These Bluetooth interactions represent short range proximity information to stationary beacons which can provide finer-grain resolution information (on the order of between 10-100 meters depending on the device) that are ubiquitous around the MIT campus. In addition, many other individuals that the subjects interacted with had similar Bluetooth-enabled mobile phones, which allow us to uniquely identify these individuals when they come in proximity to each other. According to a forecast by the International Data Corp., nearly 80% of new mobile phones sold will have Bluetooth capability by 2006. This allows us to track the social interaction patterns of a subject over time.

## **The Lifescape Dataset**

Between the behavioral data from the armband sensor, mobile phone data, and annotation information, we have managed to collect a very complete, continuous, quantifiable, and objective record of the subjects' behavior as well as detailed annotation of activities and interactions over a very long-term period of time. In total, the Lifescape Dataset consists of two and a half years (over 22,000 continuous hours) of long-term physiological/behavioral data of three subjects going about their everyday lives. The subjects consist of three graduate students from the Human

Dynamics Group at the MIT Media Laboratory, including the author. The two principal subjects participated in the study for about a year and 2 months, while the third subject yielded approximately 2 months of continuous data.

The following sections details the modeling and results primarily performed on the two principal subjects of the study.

### 8.3 Lifescape Behavior Model

We wish to be able to demonstrate that we can accurately predict the high-level behavior of individuals by looking at patterns within their past behavior as manifested in their movement and physiology. The question is how to choose the appropriate level of behavior to model. Since activities of most adults are centered around the contexts/locations such as the home and work, we decide that this is exactly the level of behavior that is most appropriate to model. Thus, we attempt to model a person's long-term patterns of behavior of moving between activities at work, home, and a catch-all 'other' which includes outside recreational activities. This captures high-level trends in behavior while black-boxing away the details of the exact activities in question, which are so varied that it would be intractable to try to identify and discriminate between them all. Cell tower IDs from the mobile phone logs are manually assigned to each of the locations of interest so we have a definitive timestamp of when the individual moves between these different high-level contexts.

Given the obvious time-dependence dynamics of human behavior, we naturally gravitate toward a Hidden Markov Model in order to analyze the data. Specifically speaking to long-term trending, HMMs as a modeling technique has been successfully used to model high-level human behavior in the past [Wren & Pentland 1999, Clarkson 2002, Eagle 2005].

As described in Section 4.10, a HMM requires a number of specifications for the model. The hidden states that we choose to model then represents the states of 'work', 'home', and 'other' ( $Q \in \{home, work, other\}$ ). Observations in the model initially are simply the collected behavioral data, such as movement or heat flux. The dataset is first preprocessed by averaging all the sensor values into minute-long chunks. In order to make the analysis more amenable, we choose to discretize the observations and inputs instead of leaving them as continuous values. A

number of different discretizations were attempted, and eventually a level around fifty discrete levels for each sensed behavioral observation.

Early experiments using just the mean values of physiology-based observation combinations with a standard HMM structure lead to very low accuracies that do not exceed the low 60% range. In order to improve accuracies, we note that most typical human behavior is heavily dependent on the normal cycles of societal activity. Intuitively, we would expect that time, specifically hour of the day and perhaps whether the day was a weekend or not, would have a clear effect on the behavior of an individual. . In fact, we can see the dependence of time by simply performing an FFT on the behavioral data observation set and looking at the resulting periodogram, which has a large spike at the frequencies corresponding to the 24-hour period.

In order to model these effects on behavior, we turn to an Input-Output HMM, whereby the states and output can be conditioned upon some additional inputs, such as time of day [Bengio & Frasconi 1995]. We can train the model using a generalized version of the EM algorithm discussed in Section 4.10. By conditioning on time, we also allow the model to account for missing data that sporadically exists in the dataset. The dataset naturally contains small blocks of data, either sensed or annotated, that are missing due to missed collections, corrupted files, and forgotten annotations.

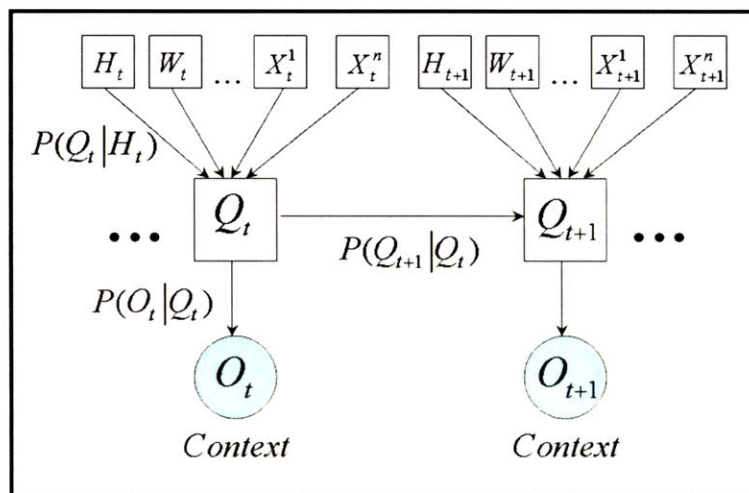


Figure 45: A Conditioned Hidden Markov Model for high-level behavior modeling. Inputs for conditioning on hour of day (H) and weekday/weekend (W) as well as behavioral inputs. We explore the effects of various inputs such as behavioral measures, physiology features, and time on the accuracies of the models in predicting the high-level behavior of an individual.

To explicitly account for this time dependence, we can condition our basic HMM on both the hour of day ( $T \in \{0, 1, 2, \dots, 23\}$ ) as well as weekday or weekend ( $W \in \{0, 1\}$ ). The model has the following graphical representation: The model was created to be flexible to allow the incorporation of additional behavioral observation vectors of interest, such movement patterns, changes in metabolism from heat flux/temperature changes, and even higher-level observations that can be derived from the behavioral data such as sleeping and activity patterns.

## 8.4 Lifescape Results

Using our conditioned HMM model, we try a variety of behavioral observation combinations to see which performs the best. The model is very flexible, in order to accommodate any number of conditioning inputs. We try both the post-processed averaged observations for the raw accelerometer (both longitudinal and transverse), heat flux, temperature, and skin conductance, as well as behavioral measures such as sedentary activity, sleep, and physical activity features. In addition, energy and entropy features of the raw sensor data were also tested on the model. We find that after training our conditioned HMM model with one and one-half months of training data, we are able to achieve relatively high levels of accuracies. Baseline distributions of times spent in home, work, and other is 46%, 38%, 16%, respectively.

By conditioning on time, we improve the results dramatically from 60% to about 73% accuracy in predicting high-level activity state. This is intuitive, as a person's behavior is highly dependent on the time of day (e.g. they normally wake up, go and leave work, and sleep at particular times). It turns out the time conditioning on hour has a great deal of effect, whereas conditioning on weekday does not. This could be due to the fact that the behavior of the subjects did not change based on the day of the week (i.e. as a student, the subject would come into lab and work as a normal day regardless of whether it was a weekday or weekend). This is perhaps idiosyncratic to the specific study population that was chosen (graduate students). It is expected that a person with a standard Monday through Friday job would show more separation of behavior between weekday and weekend, and would therefore have their modeling results improved with a weekday/weekend input parameter.



The best results came from using both channels of the raw accelerometers, with an accuracy of 73% percent. Heat flux with one accelerometer channel yielded 69% accuracies on the test data. Results with energy features were consistently a few points less than the raw values, with 71% and 65% accuracies for the two channel accelerometer and one channel accelerometer/heat flux observations, respectively.

<b>Model Inputs</b>	<b>Accuracy</b>
2-dimensional accelerometer (mean value) conditioned on H/W	73%
heat flux/ accelerometer (mean value)	69%
2-dimensional accelerometer (energy) conditioned on H/W	71%
heat flux/ accelerometer (energy) conditioned on H/W	65%
2-dim accelerometer conditioned on H/sleep	72%
2-dim accelerometer conditioned on H/activity	74%

**Table 18: Lifescape behavior modeling results using a conditioned three-state HMM corresponding to home/work/other activity classes with a variety of input feature combinations. The best performing model has time of day inputs as well as and both longitudinal and transverse accelerometer channels.**

We also attempted to condition the model by including behavioral measures using our activity-based behavior context classifiers developed in Section 5.1. Thus, instead of solely relying on time, we can accurately determine when a person sleeps and when they are active. In particular, activity is typically a reliable signal of a context switch; for example the classifier was specifically tailored to only trigger on significant sustained movement of over 15 seconds. This allows us to catch walking (which is a good indicator that a person is moving between contexts) while ignoring sporadic movements such as bumping the accelerometer into things or shifting positions temporarily. These behavioral inputs can effectively act as triggers to the Markov model to switch contextual states. In our experiments, sleep activity does not improve accuracies (as the

mutual information between time and sleep is probably very high). By adding activity as an input, we are able to slightly improve prediction accuracies to around 74%.

## 8.5 Lifescope Discussion

We have only attempted to do a very basic modeling of high-level behavior to demonstrate the feasibility of using basic physiological and behavioral sensing data that can be obtained through monitoring systems such as LiveNet. In this chapter, we demonstrate that even with the most rudimentary sensing technologies, we can predict the high-level patterns of activity of an individual to around 74% accuracy. Furthermore, the model can use very coarse grain sensor information; for our models we have down-sampled our acquired sensor data to one sample a minute using a discretization of between 5-6 bits of information.

At these low sampling resolutions and frequencies, the hardware necessary to implement a system that can be used for the type of behavior modeling we have attempted is very minimal. A watch integrated with a low-cost accelerometer with a small solid-state flash memory would be all that is necessary for the storage and sensing requirements of a device that has started to learn and predict high-level behavior to a reasonable degree of accuracy.

It is worth pointing out that it is not always important to predict with great accuracy what a person is going to do to enable useful context-aware interaction. It is often enough to give the system the equivalent of Spider Man's 'spidey sense' or a mother's intuition that something is wrong. Even if one cannot predict a person's every move throughout the day, there are certain high-level behavioral features that can be obtained from long-term trending that can result in statistical notions of what is considered normal behavior that can have large predictive power.

For example, consider the actions of Grandma Betsy, who lives alone. If she usually gets up at 8 A.M. in the morning to start her morning routine, and today, she has stayed in bed past 11 A.M., this is most likely a warning flag that should be checked out by caregivers of family. In this specific case, even a simplistic system with no learning component that simply counts wake and sleep-time statistics would have merit as a proactive system for being able to detect changes in behavior of importance.

Our model generalizes this type of behavior recognition in a very powerful way, beyond simple time and movement statistics, and allows use in exploring human behavior based on a variety of physiological, behavioral, and any other contextual cues of interest. The model can also adapt to arbitrary time-dependent changes in behavior instead of just relying on aggregate statistical measures. Furthermore, by using a well-understood methodology, we can leverage techniques like the Viterbi algorithm (discussed in Section 4.10) to find the most probable sequence of hidden states taken by the model to explain a given sequence of observations/behavior contexts. This can provide insight into how our generative model performs under different conditions so we can make further refinements to improve the model, in contrast to discriminative techniques such as neural networks which are black box in nature.

The dataset that was collected also includes very fine-grain diary information of context changes that go beyond the home/work/other that was attempted thus far. At the time of writing, the only readily-accessible and formatted context information was extracted from the cell-phone cell-tower ID and Bluetooth data logs. We are working on having these diaries transcribed and assigned with codes such that we can perform the same modeling on a wider set of behavioral contexts, including going to various other locations, performing activities, meeting with friends, etc. At this point, almost all of the subject diary information has been manually transcribed into activity, location, and interaction codes with verified timestamps, so the annotations can now be accurately synchronized to the underlying behavioral data for analysis. This analysis can be sliced in any number of dimensions across a variety of granularities.

We can also conduct our modeling on a higher granularity, and include more basic activities than the broad-based contexts of things like ‘home’ or ‘work’ activities. The basic activity classifiers developed in Chapter 5 such as working at the office, at home watching TV, standing, lying down) and associated transition activities (walking, running, going up and down stairs, riding the car) would be a good set of activities that can further add resolution to the models that were developed in this chapter.

The model framework that was developed can easily be extensible to accommodate a variety of other contexts by simply changing the number of hidden states in the model and appropriately training the model on the new set of observable contexts. Another potential extension would be to fully generalize our conditioned HMM to allow the outputs to be conditioned with inputs (i.e. to

condition the emission probabilities) the same way we condition the state transition probabilities. This would generalize the Lifescape model into a full input-output HMM model. We can use the same EM algorithm engine to learn all the necessary parameters of this generalized model.

Another interesting technique that may be applicable in this domain is by modeling activities in an online fashion using HMMs and applying selection procedure to discover the most probable model for capturing different dynamics [Sebastiani et al. 1999]. Each distinct activity of an individual can be characterized by a distinct Markov model, with the parameters of each activity learned in an online fashion. When incoming data is seen to poorly fit the evolving model of an activity, it is interpreted as a new activity and the data can be used to start fitting a new model. A technique such as this might be a more applicable model to identify activity regions to fit the Lifescape data, as there isn't much similarity between the physiologic states of performing a variety of activities such as sleeping, watching TV, and cooking, all of which get lumped into a single state 'home' in the analysis done so far.

There is also potential in exploring how behavioral context and multimodal physiology interrelate over long periods of time. One can find the repeating patterns of behavior and how they correlate to, and are influenced by, subtle physiological cues. For example, we can explore how thermoregulation cycles vary in conjunction with the movement patterns, and how this is correlated to activity through a person's everyday behavior using the Influence Model, a coupled-HMM formulation that has been shown to be very effective at modeling things as diverse as power grids and conversation dynamics [Asavathiratham 2000, Basu 2001]. The Influence Model models each individual data stream, or agent, of interest with an independent Markov chain. Interaction between these data streams can be learned through the "influence" parameters that quantify the degree to which each agent affects the behavior of the other. The benefit of this model is that it provides a way to quantify inter-agent interaction while still being computationally tractable and scaleable with the number of agents being modeled. In fact, there is work being done now to optimize Influence Model code (currently available only in Matlab) to be able to run locally on Zaurus embedded hardware.

## Chapter 9      Future Directions

In this chapter, we discuss potential future directions in which to extend the existing corpus of research that this thesis demonstrates. In the sections below, we discuss how the research performed in this thesis can be extended to develop real-time feedback systems to enable rich proactive healthcare applications in the domain of clinical diagnosis and eldercare. We also describe how the existing LiveNet platform can be used to develop distributed healthcare systems that allow for social support.

### 9.1    Real-Time Feedback Applications

The research that was presented has focused on the monitoring and modeling of individual human physiology and behavior in a variety of settings. One aspect that has only been lightly explored is how to build practical systems that do not just monitor, but provide feedback to the user. The ultimate goal of this type of research is to close the loop by using the inference and classification results of such models to provide feedback to the user in a meaningful manner.

We touched upon some of these feedback applications by using LiveNet to prototype some real-time classifiers, such as the context classifiers and soldier monitors discussed in Chapters 5 and 6, respectively. Likewise, while we did not implement a real-time classification system base on the depression research in Chapter 7, it would not be hard to create a simple non-invasive depression classifier. Even if such a classifier is not highly accurate, this type of application could still potentially have very positive consequences, such as perhaps saving a person from suicide.

The Lifescape study demonstrates that we can use very non-invasive sensing to predict high-level behavior. By integrating the accelerometer into an existing device that people are comfortable carrying around (for example, a cell phone), we can significantly lower the bar for developing a practical activity classification system to the mass market that is completely transparent to the user.

When combined with physiological measurements such as heart rate and breathing rate, these measurements can then be collected to build a personalized profile of your body's performance and your nervous system's activation throughout your entire day, and assembled over a period of

months or years to show long-term changes in overall cardiac fitness. Computer software 'agents' (automatic computer programs) could even give you gentle reminders to keep up your routine if your activity level started to decline and make suggestions to optimize your performance.

## **9.2 Distributed Peer-to-Peer Healthcare Applications**

One very interesting area that was not explored in this thesis is the peer-to-peer sharing of health information in distributed settings. Traditionally, mobile healthcare applications are usually thought of in the context of single users, such as individualized monitoring, bio-feedback, and assistive PIM applications. There is also new opportunity in making this information explicit not only to oneself, but one's peer/social network for community support and interaction.

A platform such as LiveNet can allow us to extend this paradigm by using mobile technology to also assist and augment group collaboration and other types of social interaction. Applications of real-time distributed physiology context are numerous, enabling new forms of social interaction and communication for community-based support by peers. Because the LiveNet system is a distributed system that is capable of streaming data to arbitrary endpoints, one can imagine a variety of telemedicine applications where real-time feedback of important health state can be relayed to individuals as well as to doctors and other caregivers.

However, the analogy of sharing health information can be generalized beyond the specific critical health monitoring domain for such things as soldier systems and hospital networks to health information networking, or the sharing of long-term health and physiology information with peer groups in a transparent and meaningful manner. Associated benefits include the sense of telepresence or virtual proximity of peer groups through multimodal data streams (including audio/video, physiological, and other contextual cues) and "Peer-to-Peer" pressure for eliciting healthy behavior in groups via the explicitly laid out social support infrastructure. We can use the social network to enforce positive healthy behavior, e.g. providing instant support to encourage not smoking/eating compulsively/drinking/doing drugs when a person is vulnerable, either by automated behavior classification or even if the person explicitly makes it known.

The LiveNet platform was designed to implement distributed group applications involving streaming real-time information and context results. LiveNet systems can also be configured to

enable point-to-point real-time audio data streams, functioning as communicators that can provide instantaneous voice interaction to groups. By simply pushing a button, individuals can communicate with each other without effort, thereby strengthening the social network and peer support groups. Reducing the effort needed to communicate can significantly change group interaction dynamics, much as instant messaging revolutionized text communication relative to slower email communication. Through these instantaneous communication channels, it is possible to create the feeling of virtual proximity of social peer or family groups, despite the fact that there may be large geographical separation between individuals.

In addition to the reduced communication barriers, improved implicit awareness of the group through broadcasted psychophysiological cues is possible, creating a 'telepresence' that is capable of establishing stronger social ties within groups. For example, a group's real-time psychophysiology statistics can be aggregated, broadcast, and displayed to location-separated endpoints and used to gauge group status in the form of socio-biofeedback. These types of socio-physiometric displays have the power to radically change the dynamics and interaction between people, as well as to provide additional dimensions of information dissemination.

One of the main problems with new health technology, particularly for the technophobic elderly, is that it can be very impersonal and distracting. As such, compliance has always been one of the largest problems of using information technologies within the healthcare industry. Social group activities provide a compelling way to support and enforce compliance, potentially eliciting healthy behavior in peer groups. One example of this is DiaBetNet [Kumar & Pentland 2002], a distributed PDA-based handheld game that has been shown to help diabetic children learn to regulate their own blood sugar levels through competitive community games.

Another potentially interesting idea is to create a distributed community exercise game, such as using stationary erging/stairmaster/exercise bikes, and demonstrating that by displaying real-time health information such as heart rate (maintaining heart rate targets has been en vogue as a purported method for optimal body exercise training) or exercise output, one can increase the overall efficiency of the entire group's workout through fun competition. Likewise, similar information can be used in team settings such as in crew boats, for biofeedback and coordination to match output for optimal team performance through a race. The LiveNet platform is a perfect

platform for developing these types of community games and other support structures to elicit healthy behavior in peer groups.

The BorgViewer application is an example of a basic peer-to-peer distributed application as a demonstration of how this health networking concept can be used effectively in a practical healthcare setting. It is not hard to see how we can generalize the BorgViewer application to command and control applications, such as a platoon of soldiers who need to communicate with each other and immediately identify each individual's status, to coordinate activities and work in a cohesive manner.



## Chapter 10      Conclusions

The work in this thesis touches upon a broad spectrum of interdisciplinary topics, including practical wearable computing infrastructure, human factors research, embedded systems development and integration, non-invasive physiological and contextual sensing technologies, machine inference and modeling, distributed system and context-aware application design, and clinical research. Using the infrastructure and tools developed for the LiveNet Platform, we have been able to tackle some interesting problems in context classification, continuous monitoring, clinical pathology trending, and long-term human behavior modeling. Some of the accomplishments of the thesis are highlighted and summarized in the sections below.

### 10.1    Prototyping Technologies for Future Healthcare Applications

The essential philosophy behind this thesis is to leverage the power of combining low-cost commodity hardware and non-invasive sensing in creating powerful but practical context-aware systems. We have attempted to construct a flexible wearable sensing and monitoring platform centered around a sensor hub board called the SAK2, which allows us to leverage the power of modern mobile devices such as smart phones and PDAs while providing a flexible infrastructure to simultaneously connect a wide range of customized and commercially available sensors.

We attempt to get the most mileage out of using simple non-invasive sensing technologies to develop powerful context-aware systems that can enable these proactive healthcare applications. We show that simple minimally-invasive, noisy, non-expensive physiologic sensing features can be correlated and used as effective substitutes to more invasive health sensing for robust context classification, thereby enabling practical mobile health applications. We also develop a few general contextual classifiers for things such as activity, stress level, cold-exposure, and depression state based on simple physiologic cues that can be detected using non-invasive sensing.

By combining continuous monitoring of non-invasive physiologic and behavioral variables in a mobile multimodal sensing device, LiveNet can be used to explore domains that were previously impractical. Past limitations of not being able to attach a configurable sensor network to a mobile device has been solved by the embedded hardware developed for LiveNet. Also, by using non-

invasive sensing, we can reduce the burden on potential subjects and thereby extend the time that we can monitor them, as well as allow the monitoring to take place in non-laboratory settings. The particular Zaurus instantiation of the LiveNet platform has proven to be quite capable of long-term ambulatory monitoring.

The LiveNet platform's strengths lay in the fact that the system is highly customizable and extendable, allowing researchers the necessary infrastructure to rapidly prototype a variety of relevant healthcare applications. We have developed a wearable hardware, biosensor, and systems infrastructure that has proven to be robust and practical for long-term continuous monitoring applications. Using LiveNet, we can quickly develop customized wearable sensing systems capable of multimodal sensing and monitoring, real-time data streaming, context-aware inference, and distributed networking. The sensor data and real-time classification results from a LiveNet system can also be streamed to off-body servers for subsequent processing, trigger alarms or notify family members and caregivers, or displayed/processed by other LiveNet systems or computers connected to the data streams for complex real-time interactions.

We have also developed a suite of LiveNet applications that allow us to easily conduct a variety of large-scale data acquisition, monitoring, and real-time feedback applications. Building upon the base infrastructure of LiveNet, this practical set of tools enables us to build turn-key systems that can be used by researchers to collect data for short experiments, to continuously monitor individuals in long-term ambulatory settings, or even to develop systems capable of real-time inference, feedback, and interaction.

Given its real-time processing and streaming data capabilities, the LiveNet system allows people to receive real-time feedback from their continuously monitored and analyzed health state, as well as communicate health information with care-givers and other members of an individual's social network for support and interaction. These types of technology will allow both individuals and groups of people to manage their health and support each other more effectively.

## 10.2 Data-mining Individuals: Exploring Human Physiology and Behavior

Using the LiveNet platform, we have been able to quantitatively data-mine long-term human physiology and behavior. We demonstrate that LiveNet can be used as a practical platform to conduct clinically significant research in both laboratory and ambulatory settings, specifically in the context of health state classification and long-term continuous monitoring.

Through the studies conducted, we have been able to quantitatively explore and analyze human physiology and behavior. We have generated several novel datasets spanning a variety of interesting healthcare domains. We use the data collected to uncover how continuously-monitored physiology and immediate contextual state can be correlated to long-term health state and behavior. We demonstrate that long-term monitoring can allow one to track trends in human physiology in a quantitative, continuous manner that hasn't been possible in the past, given technological limitations. Using the obtained physiological and behavioral data, we have developed accurate multi-linear regression models that can trend a patient's subjective emotions and clinical outcomes through the course of medication and treatment. We also show that we use non-invasive physiology data to develop generative probabilistic graphical models that can be used to classify and predict an individual's long-term behavior and time-dependent dynamics.

We demonstrate how the non-invasive sensing can be applied to develop interesting context-aware systems. In particular, we show how motion sensing enables us to develop powerful classifiers capable of differentiating between a variety of everyday activities. We also demonstrate that we use non-invasive physiology to develop accurate classifiers that can be used to detect stress and lying in the PokerMetrics Study.

In addition, we show how we can use motion sensing to develop highly accurate real-time shivering monitors in the ARIEM Soldier Monitoring Study. We then demonstrate how we can use Hidden Markov Models to develop model the time dynamics of shivering and correlate motion to core body temperature. The resulting classification systems can triage a soldier's cold exposure state with near-perfect accuracy. Once we develop more insight into how physiology and behavior are interrelated in the domains of interest, we can start to develop practical systems for real-time classification and feedback based on this research. In this thesis, we demonstrate that

we can monitor and provide real-time feedback (such as the real-time classifier with the ARIEM Cold-Exposure Study).

In addition to the successes of using the LiveNet system for real-time context classification applications, we have shown that we can also use the system to conduct long-term monitoring studies. In Chapter 7, we discuss the MGH Depression Study, which monitors the continuous physiology of clinically depressed patients at the psychiatric ward at the Massachusetts General Hospital. We detail the protocol, and then follow up with an in-depth analysis of how the physiology sensing and behavior data can be correlated to the subjective emotion ratings of the patients. We then follow by showing how these features can be used to create multi-linear regression models that can accurately trend clinical outcomes of depression.

The ultimate goal is to enable proactive applications that take the health monitoring out of strict medical and hospital contexts. Toward this end, we initiated the Lifescape Study, where over a year and a half of long-term physiology and behavioral data was collected on a number of subjects. We demonstrate how we can use conditioned Hidden Markov Models as a means of behavior parameterization and prediction. Using this model, we show how we can model the long-term behavioral dynamics of individuals going about their daily lives solely with the use of non-invasive physiological sensing. Using non-invasive and simple contextual cues, we hope to be able to demonstrate that it is possible to model a person's long-term behavior. Deviations from this behavior, which could indicate a warning sign, can then be used as a basis for intervention.

LiveNet provides a convenient platform to be able to quantify the long-term context and physiological response of individuals. This, in turn, will support the development of individualized treatment systems with real-time feedback to help promote proper behavior. The ultimate goal of the research is to eventually be able to use LiveNet for developing practical monitoring systems and therapeutic interventions in ambulatory, long-term use environments.

### **10.3 Privacy Implications and the Price of Privacy**

Long-term monitoring applications have the potential to dramatically transform the ways in which we can track and identify changes in health over time. For this to happen, though, researchers need to address a number of privacy concerns. The privacy concerns involving long-

term continuous monitoring of behavior and health are numerous. A system that saves personal health and behavior information and potentially transmits this information to other endpoints is always open to liability and privacy issues. Utmost care must be made to ensure that sensitive personal health information is properly protected and controlled.

We can provide some basic security checks to mitigate some of the risk. At a physical level, the LiveNet platform is capable of SSL encryption that can be used to protect all data streams leaving the local device. We have also tried to protect the privacy of the subjects in the studies by only recording features which have no useable information by themselves. For example, by only recording high-level statistical features of the basic properties of speech such as pitch/magnitude, we do not save any identifiable speech content. The speech features we extract and store cannot be traced back to the subject, or be used to reconstruct the conversational content. Consequently, sensitive information in the audio is rendered safe. Furthermore, we have demonstrated that in many of the relevant instances, we do not even require detailed physiology information. Aggregated physiology features, such as in the stress and depression studies, is all that is necessary in order to enable the context classifications we are interested in.

Another concern is the generally unease many people feel about having their movements and behavior tracked. However, in general, studies have found that people are usually willing to relinquish a portion of their privacy in exchange for something of (typically surprisingly small) value [Huberman et al. (2005)]. Consumers, for example, have been willing to divulge personal information, such as the names of their friends and relatives, to receive free gifts or reduced rates for a service. For the majority of people, the benefits of paying with a credit card outweigh the perceived intrusion of providing a company access to information on the location and content of each purchase.

Likewise, in many instances behavior tracking can provide information that may be beneficial to the overall long-term health of an individual. In such situations, users may be willing to give up a portion of their privacy in exchange for the benefits the monitoring provides. Also, in general, much of the behavior monitoring can be accomplished in a local manner. Thus, the personal health and behavior information that is derived does not reside anywhere except on the device that the user has direct physical control of at all times. In cases where health information must be transmitted, centralized (instead of peer-to-peer) control helps ensure that people share only the

information that they want to share. People should be able to mediate who has access to certain data. A user might, for example, specify that certain pieces of information be shared only with family or authorized healthcare providers.

Whether it is through encryption, health information policies, or simply limiting interactions to users within a friends-of-friends trust-network, it is clear that there is a need to make privacy-protecting tools in order to maintain a user's right and expectation of privacy. These privacy concerns will need to be addressed in tandem to the development of practical monitoring applications that use sensitive health information.

## **10.4 Limitations to LiveNet**

Hopefully we have demonstrated that we can leverage commodity hardware and simple non-invasive sensing technologies for long-term monitoring applications and real-time context-classification systems. At essence, the LiveNet platform was conceived as a set of practical, modular, extensible, and reliable tools to enable clinical research. While we have been able to develop practical wearable sensing systems that are relatively less of a burden and more affordable than most existing health monitoring technologies, the LiveNet platform is still a ways off from being completely transparent to the user. We have made conscious trade-offs for improved modularity, extensibility, and affordability at the cost of some optimizations in form factor, integration, and interface. An analogy is that we have created a platform that is a swiss-army knife capable of doing many different things, which may or may not be ideal for use in real-world settings. Thus, while LiveNet is a good platform for rapidly prototyping applications and enabling large-scale rollouts for use in clinical studies, it is still a tool that must be further refined before it is applicable for practical end-user systems.

Also, by focusing on commodity hardware, we are essentially limited to devices that may be inherently unreliable and limited in functionality (such as I/O). While we have demonstrated that we can develop relatively robust systems based on low-cost mobile devices for use in long-term settings, there are still reliability issues that may make the platform inappropriate in some medical settings. Specifically, for devices which need FDA approval for use in settings such as critical health monitoring, the price of failure can literally cost lives and hence reliability and accuracy is of paramount importance. Admittedly, our philosophy tends toward developing applications

which provide high levels of benefit without needing to be completely reliable or accurate and where the cost of failure is not particularly onerous.

While the LiveNet system has been designed with comfort and ease of use in mind, it is noted that it may not yet be practical and burdenless for the general population to wear PDA-based systems for long periods of time. However, it is important to point out that the LiveNet system is agnostic to the underlying hardware, and within a few generations (i.e., a couple of years), new much smaller, faster, integrated-functionally cellphones which a person could conceivably slip into their pocket and forget about will exist. The LiveNet system therefore demonstrates a system that is flexible enough to eventually evolve into something that will be appropriate for practical future long-term health monitoring applications.

Finally, since our focus is on non-invasive sensing, there are of course certain limitations in terms of resolution, reliability, and measurement capabilities relative to more invasive sensing. Some types of relevant information, such as blood sugar levels in diabetics, are still not able to be accurately derived through non-invasive means given existing technology. Thus, applications in these domains must still rely on more invasive forms of sensing, at the cost of convenience and portability.

## **10.5 Glimpsing into the Future of Healthcare**

We are about to enter a truly exciting period in time in history. Technology is rapidly approaching the inflection point where we can truly start to monitor individuals on a continuous basis in a completely transparent way. This has the potential to revolutionize research in a variety of academic disciplines, including medicine, psychology, sociology, behavioral economics, and ethnography, to name a few.

Minimally invasive sensing expands the relevance of monitoring systems outside the clinical context in terms of the potential to being able to monitor everyday behavior. As technology shrinks and becomes less and less unwieldy and invasive, it is feasible that we are on the horizon of a world where it may be commonplace for people to wear long-term health monitoring systems even when they are not sick. The point is to move to a more proactive healthcare paradigm in which computer-assisted technology tracks personal health trends and non-healthy behaviors are

monitored and corrected before they become chronic (that on top of the standard critical health monitoring applications that people focus on today).

It turns out that medicine has not focused on the science of physiological changes over the long term, primarily because in the past there was no way of monitoring continuously throughout a person's life. Instead, long-term physiological monitoring currently consists of spot checks from physical exams, which occur on the order of months if not years. Thus, it is only within critical health domains that people have explicit notions of their semi-continuous physiology, such as diabetics who need to know their blood sugar levels to properly regulate their health maintenance.

This mentality will change with time because of technology improvements that will make an individual's own health state more explicit. This is imperative for more proactive health management paradigms. Currently, people do not have any technologies that can help them monitor and track their health state, thus relying on hospitals when they become sick to fix a problem that perhaps could have been fixed had more health information been available. With improved self-awareness of their own physiology, they can be more proactive to maintain their health before they become sick and have to go to the hospital.

Progress in terms of understanding human physiology and behavior will result from the fact that long-term trends in human physiology can be explored in detail. This has some profound ramifications that will reshape our understanding of the human body and human behavior. Such advances include tracking the development and evolution of diseases, monitoring changes in physiology as people grow older, comparing physiology across different populations (gender, ethnicity, etc), and even knowing characteristic physiology patterns of people who are *healthy* (this last example is particularly important when it is necessary as a diagnostic methodology designed to quantitatively define abnormal behavior).

From continuous monitoring, a very fine granularity of quantitative data can be obtained, in contrast to the surveys and history-taking which have been the mainstays of long-term studies and health interventions in the past. The next step is to be able to detect repeating trends in complex human physiology and behavior by analyzing the patterns in data collected. The developed models and trends can then be incorporated into proactive applications such as early detection and decision guidance systems, rehabilitation, and personal health status informatics. These proactive



systems with real-time feedback can be used to promote healthy, preventative behavior and help people comply with them.

With the development of ambulatory systems capable of real-time feedback, we can move to a paradigm of a more personalized medication and rehabilitation scheduling based on measured, quantitative physiological symptoms and behavioral responses, not based only on time scheduled approximations as the current practice. The effects of medication and rehabilitation treatment can be logged and recorded quantitatively and compared to changes in physiology, eventually with a real-time, closed-loop monitoring and drug delivery system that can track the effects of individualized treatment over time.

In general, with real-time sensing and processing capabilities, it will become possible to effectively reduce the time delay in processing and receiving health feedback. This is particularly true when the doctor can be removed from the equation, such as by using real-time diagnostic systems that can provide effectively instantaneous classification on health state and context. This will hopefully reduce the need for direct physician interaction and reduce the burden on the healthcare system. At the same time, by providing more relevant feedback and information, we can empower individuals to be more in tune with their own health and allow them to make more informed decisions for themselves.

## References

- [Abowd & Mynatt 2000] G. Abowd & E. Mynatt (2000) "Charting Past, Present, and Future Research in Ubiquitous Computing", *ACM Transactions on Computer-Human Interaction*, March, 2000
- [Abowd 2003] G. Abowd (2003) "Improving Outlooks for Aging Baby Boomers", Testimony before the Senate on Special Aging, 107th Congress, 2003
- [Adamson et al. 2004] P. Adamson, A. Smith, W. Abraham, K. Kleckner, R. Stadler, A. Shih, and M. Rhodes (2004) "Continuous Autonomic Assessment in Patients with Symptomatic Heart Failure: Prognostic Value of Heart Rate Variability Measured by an Implanted Cardiac Resynchronization Device", *Circulation*, Aug. 2004
- [Alpert et al. 1997] J. Alpert, L. Uebelacker, N. McLean, A. Nierenberg, J. Pava, J. Worthington III, J. Tedlow, J. Rosenbaum, and M. Fava (1997) "Social Phobia, Avoidant Personality Disorder and Atypical Depression: Co-Occurrence and Clinical Implications", *Psychological Medicine*, 27: 627-633
- [Ambady & Rosenthal 1992] N. Ambady & R. Rosenthal (1992) "Thin Slices of Expressive Behaviour as Predictors of Interpersonal Consequences: A Meta-Analysis", *Psychological Bulletin*, 111(2): 256-274
- [Andersson et al. 1999] S. Andersson, J. Krogstad, and A. Finset (1999) "Apathy and Depressed Mood in Acquired Brain Damage: Relationship to Lesion Localization and Psychophysiological Reactivity", *Psychological Medicine*
- [Argyle 1987] M. Argyle (1987) *Bodily Communication*, Methuen
- [Arruda et al. 1996] J. Arruda, R. Stern, and S. Legendre (1996) "Assessment of Mood State in Patients Undergoing Electroconvulsive Therapy: The Utility of Visual Analog Mood Scales Developed for Cognitively Impaired Patients", *Convuls Ther.*, 12(4):207-12
- [Asada et al. 2003] H. Asada, P. Shaltis, A. Reisner, S. Rhee (2003) "Mobile Monitoring with Wearable Photoplethysmographic Biosensors", *IEEE Engineering in Medicine and Biology Magazine*
- [Asavathiratham 2000] C. Asavathiratham (2000) "The Influence Model: A Tractable Representation for the Dynamics of Networked Markov Chains", Ph.D. Thesis, MIT Department of Electrical Engineering and Computer Science
- [Balogh et al. 1993] S. Balogh, D. Fitzpatrick, S. Hendricks, and S. Paige (1993) "Increases in Heart Rate Variability with Successful Treatment in Patients with Major Depressive Disorder", *Psychopharmacology Bulletin*, 29(2):201-206

**[Bao 2003]** L. Bao (2003) “Physical Activity Recognition from Acceleration Data under Semi-Naturalistic Conditions”, M.Eng. Thesis, MIT Dept. Electrical Engineering and Computer Science

**[Bashshur et al. 1975]** R. Bashshur, P. Armstrong, and Z. Youssef (1975) *Telemedicine: Explorations in the Use of Telecommunications in Health Care*, Charles C. Thomas, Springfield, IL

**[Bashshur 1980]** R. Bashshur (1980) “Technology Serves the People: The Story of a Cooperative Telemedicine Project by NASA, the Indian Health Service and the Papgo People”, Superintendent of Documents, US Government Printing Office, 1980

**[Bashshur et al. 1997]** R. Bashshur, J. Sanders, and G. Shannon (1997) *Telemedicine: Theory and Practice*, Charles C Thomas, Springfield, IL

**[Basu et al. 2001]** S. Basu, T. Choudhury, B. Clarkson, and A. Pentland (2001) “Learning Human Interactions with the Influence Model”, Technical Report, Human Dynamics Group

**[Basu 2002]** S. Basu, (2002) Conversational Scene Analysis, Doctoral Thesis, MIT Dept. of Electrical Engineering and Computer Science

**[Bell et al. 1992]** D. Bell, P. Tikuisis, and I. Jacobs (1992) “Relative Intensity of Muscular Contraction During Shivering”, *Journal of Applied Physiology*, Vol. 72

**[Bengio & Frasconi 1995]** Y. Bengio & P. Frasconi (1995) “Input/Output HMMs for Sequence Processing”, *IEEE Transactions on Neural Networks*, 7(5):1231-1249

**[Bilmes 1997]** J. Bilmes (1997) “A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models” Technical Report # ICSI-TR-97-021, University of California, Berkeley

**[Binder et al. 1992]** T. Binder, B. Frey, G. Porenta, G. Heinz, M. Wutte, G. Kreiner, H. Gossinger, H. Schmidinger, R. Pacher, and H. Weber (1992) “Prognostic Value of Heart Rate Variability in Patients Awaiting Cardiac Transplantation”, *PACE*, 15:2215-2220

**[BodyMedia]** BodyMedia Corporation, <http://www.bodymedia.com>

**[Bolis & Licinio 1999]** C. Bolis & J. Licinio, Editors (1999) “The Autonomic Nervous System”, World Health Organization, Geneva

**[Bonato 2003]** P. Bonato (2003) “Wearable Sensors/Systems and Their Impact on Biomedical Engineering”, *IEEE Engineering of Medicine and Biology*

**[Brady & Kendall 1992]** E. Brady & P. Kendall (1992) “Comorbidity of Anxiety and Depression in Children and Adolescents”, *Psychol Bull.*, 111(2):244-55

- [Brodaty et al. 2001]** H. Brodaty, D. Berle, I. Hickie, and C. Mason (2001) “‘Side Effects’ of ECT are Mainly Depressive Phenomena and are Independent of Age”, *Journal of Affective Disorders*, 66:237-245
- [Brown et al. 1997]** P. Brown, J. Bovey, and X. Chen (1997) “Context-Aware Applications: From the Laboratory to the Marketplace”, *IEEE Personal Communications*, October
- [Buchanan et al. 1985]** S. Buchanan, J. Valentine, and D. Powell (1985) “Autonomic Responses are Elicited by Electrical Stimulation of the Medial but not Lateral Frontal Cortex in Rabbits”, *Behavioral Brain Research*
- [Bussmann et al. 2001]** J. Bussmann, W. Martens, J. Tulen, F. Schasfoort, H. van den Berg-Emons, and H. Stam (2001) “Measuring Daily Behavior Using Ambulatory Accelerometry: The Activity Monitor” *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356
- [Cacioppo & Tassinari 1990]** J. Cacioppo and L. Tassinari (1990) “Inferring Psychological Significance from Physiological Signals”, *American Psychologist*, 45(1):16-28
- [Cacioppo et al. 2000]** J. Cacioppo, L. Tassinari, and G. Berntson (2000) *Handbook of Psychophysiology*, Cambridge University Press
- [Camacho et al. 1991]** T. Camacho, R. Roberts, N. Lazarus, G. Kaplan, and R. Cohen (1991) “Physical Activity and Depression: Evidence from the Alameda County Study”, *American Journal of Epidemiology*, 134(2):220-231
- [Cammann & Michel 2002]** H. Cammann & J. Michel (2002) “How to Avoid Misinterpretation of Heart Rate Variability Power Spectra”, *Computer Methods Programs Biomed.*
- [Carney et al. 1988]** R. Carney, M. Rich, A. teVelde, J. Saini, K. Clark, K. Freedland (1988) “The Relationship Between Heart Rate, Heart Rate Variability and Depression in Patients with Coronary Artery Disease”, *Journal of Psychosomatic Research*, 32(2):159-164
- [Caro 2003]** M. Caro (2003) *Caro’s Book of Poker Tells*, Cardoza Publishing
- [Carpeggiani et al. 2004]** C. Carpeggiani, A. L’Abbate, P. Landi, C. Michelassi, M. Raciti, A. Macerata., and M. Emdin (2004) “Early Assessment of Heart Rate Variability is Predicting of In-Hospital Death and Major Complications after Acute Myocardial Infarction”, *International Journal of Cardiology*
- [Cartwright et al. 2003]** R. Cartwright, E. Baehr, J. Kirkby, S. Pandi-Perumal, J. Kabat (2003) “REM Sleep Reduction, Mood Regulation and Remission in Untreated Depression”, *Psychiatry Res.*, 121(2):159-67
- [Cassell & Bickmore 2003]** J. Cassell & T. Bickmore (2003) “Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents”, *User Modeling and User-Adapted Interaction*, 13(1-2): 89-132

- [**Chambers & Allen 2002**] A. Chambers & J. Allen (2002) “Vagal Tone as an Indicator of Treatment Response in Major Depression”, *Psychophysiology*
- [**Chambers et al. 2002**] G. Chambers, S. Venkatesh, G. West, and H. Bui (2002) “Hierarchical Recognition of Intentional Human Gestures for Sports Video Annotation”, *International Conference on Pattern Recognition*, IEEE Press
- [**Chartrand & Bargh 1999**] T. Chartrand & J. Bargh (1999) “The Chameleon Effect: The Perception-Behavior Link and Social Interaction”, *Journal of Personality and Social Psychology*, 76(6): 893-910
- [**Choudhury & Pentland 2003**] T. Choudhury & A. Pentland (2003) “Modeling Face-to-Face Communication Using the Sociometer”, *IEEE International Conference of Ubiquitous Computing*
- [**Choudhury 2003**] T. Choudhury (2003) “Sensing and Modeling Human Networks”, Doctoral Thesis, MIT Dept. of Electrical Engineering and Computer Science
- [**Clark & Watson 1991**] L. Clark & D. Watson (1991) “Tripartite Model of Anxiety and Depression: Psychometric Evidence and Taxonomic Implications”, *J Abnorm Psychol.* 100(3):316-36
- [**Clarkson et al. 2000**] B. Clarkson, K. Mase, and A. Pentland (2000) “Recognizing User Context via Wearable Sensors”, *IEEE International Symposium on Wearable Computers*, pp. 69–76.
- [**Clarkson 2002**] B. Clarkson (2002) “Life Patterns: Structure from Wearable Sensors”, Ph.D. Thesis, MIT Media Arts and Sciences
- [**Dawson et al. 1977**] M. Dawson, A. Schell, and J. Catania (1977) “Autonomic Correlates of Depression and Clinical Improvement Following Electroconvulsive Shock Therapy”, *Psychophysiology*
- [**Dawson et al. 1985**] M. Dawson, et al. (1985) “Diagnostic Utility of Autonomic Measures for Major Depressive Disorders”, *Psychiatry Research*
- [**Dawson et al. 1990**] M. Dawson, A. Schell, and D. Fillion (1990) “The Electrodermal System” In Cacioppo and Tassinari, editors, *Principles of Psychophysiology: Physical, social and inferential elements*, pp. 295-324. Cambridge University Press, Cambridge
- [**DeJohn et al. 1992**] C. DeJohn, M. Marr, E. Molina, and A. McCardie, (1992) “A Computer-Based Visual Analog Scale”, Technical Report # A251852, Naval Aerospace Medical Research Labs, Pensacola, FL.
- [**Devanand et al. 1995**] D. Devanand, L. Fitzsimons, J. Prudic, and H. Sackeim (1995) “Subjective Side Effects During Electroconvulsive Therapy”, *Convuls Ther.*, 11(4):232-240
- [**DeVaul et al. 2001**] R. DeVaul, S. Schwartz, and A. Pentland (2001) “MIThril: Context-Aware Computing for Daily Life”, Whitepaper, MIT Media Laboratory

[DeVaul 2002] R. DeVaul (2002) "Gaussian Mixture Model Tutorial", <http://www.media.mit.edu/wearables/mithril/BNT/mixtureBNT.txt>

[DeVaul & Pentland 2003] R. DeVaul & A. Pentland (2003) "The MIThril Real-Time Context Engine and Activity Classification", Technical Report, MIT Media Laboratory

[DeVaul et al. 2003] R. DeVaul, M. Sung, J. Gips, and A. Pentland (2003) "MIThril 2003: Applications and Architecture", *IEEE International Symposium of Wearable Computers*

[Digital Angel] Digital Angel Corporation, [www.digitalangelcorp.com](http://www.digitalangelcorp.com)

[Dougherty & Rauch 1997] D. Dougherty & S. Rauch (1997) "Neuroimaging and Neurobiological Models of Depression", *Harvard Review of Psychiatry*

[Dougherty & Rauch 2001] D. Dougherty & S. Rauch (2001) *Psychiatric Neuroimaging Research: Contemporary Strategies*, American Psychiatric Publishing, Washington DC

[Drevets 2001] W. Drevets (2001) "Neuroimaging and Neuropathological Studies of Depression: Implications for the Cognitive-Emotional Features of Mood Disorders", *Current Opinion in Neurobiology*

[DSM 1994] DSM-IV, T.F.o., (1994) "Diagnostic and Statistical Manual of Mental Disorders", *American Psychiatric Society*, Washington, D.C.

[Duda & Hart 2000] R. Duda, P. Hart, and D. Stork (2000) *Pattern Classification*, John Wiley & Sons

[Eagle & Pentland 2003] N. Eagle & A. Pentland, "Social Network Computing", *Conference on Ubiquitous Computing*

[Eagle & Pentland 2004] N. Eagle & A. Pentland (2004) "Social Serendipity: Proximity Sensing and Cueing" *IEEE Pervasive Special Issue on Smart Phones*

[Eagle 2005] N. Eagle, (2005) Machine Perception and Learning of Complex Social Systems, Ph.D. Thesis, MIT Media Arts and Sciences

[Ekman & Friesen 1977] Ekman, P. and Friesen, W. (1977) *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, CA

[Eisenberg et al. 1998] D. Eisenberg, R. Davis, S. Ettner, S. Appel, S. Wilkey, M. Van Rompay, and R. Kessler (1998) "Trends in Alternative Medicine Use in the United States, 1990-1997: Results of a Follow-up National Survey," *JAMA*

[Elledge 2003] C. Elledge (2003) "Wearable Multiple Sensor Acquisition Board", M. Eng. Thesis, MIT Media Art and Sciences

[Endicott et al. 1993] J. Endicott, et al. (1993) "Quality of Life Enjoyment and Satisfaction Questionnaire: A New Measure", *Psychopharmacology Bulletin*

[Fauchet et. al. 2004] P. Fauchet, A. Pentland, and C. Horwitz (2004) "Letting the Home Interface with the Health Care System: New Paradigms for Consumers and Providers", Thought Leaders' Workshop White Paper, University of Rochester Center for Future Health

[FDA 2001] Federal Food and Drug Administration (2001) "FDA 510(k) Summary for LifeShirt System with Vivologic Analysis Software", <http://www.fda.gov/cdrh/pdf/k011903.pdf>

[FDA 2003] Federal Food and Drug Administration (2003) "Medical Devices: Cardiovascular Devices: Reclassification of the Arrhythmia Detector and Alarm. Final Rule", *Fed Regist.*

[FitSense] FitSense Corporation, <http://www.fitsense.com>

[Fogarty 2004] J. Fogarty (2004) "Connection versus Calming: Interruptibility, Presence, and Availability", Position paper, Workshop on Forecasting Presence and Availability, *CHI*

[Foerster et al. 1999] F. Foerster, M. Smeja, and J. Fahrenberg (1999) "Detection of Posture and Motion by Accelerometry: A Validation in Ambulatory Monitoring", *Computers in Human Behavior*

[Forney 1973] G. Forney (1973) "The Viterbi Algorithm", *Proceedings of the IEEE* 61(3):268-278

[Frasure-Smith 1991] N. Frasure-Smith (1991) "In-hospital Symptoms of Psychological Stress as Predictors of Long-Term Outcome After Acute Myocardial Infarction in Men", *Am J Cardiol*, 67:121-127

[Frigo & Johnson 2005] M. Frigo & S. Johnson (2005) "The Design and Implementation of FFTW3", *Proceedings of the IEEE*, 93(2):216-2312

[Frijda 1986] N. Frijda (1986) "The Emotions, Physiology of Emotion", pp. 124-175, *Studies in Emotion and Social Interaction*, Cambridge University Press, Cambridge

[Fuller 1977] G. Fuller (1977) "Biofeedback Methods and Procedures in Clinical Practice", Biofeedback Institute of San Francisco

[Gerasimov 2003] V. Gerasimov (2003) "Every Sign of Life", Ph.D. thesis, MIT Media Arts and Sciences

[Gift 1989] A. Gift (1989) "Visual Analog Scales: Measurement of Subjective Phenomenon", *Nursing Research*, 38(5):286-288

[Glass 1985] R. Glass (1985) "Electroconvulsive Therapy: Time to Bring It Out of the Shadows", *JAMA*, 285:1346-1348, Electroconvulsive Therapy, NIH Consensus Statement

[Gladwell 2000] M. Gladwell (2000) *The Tipping Point: How Little Things Can Make a Big Difference*, Little Brown and Company, New York

[Gladwell 2005] M. Gladwell (2005) *Blink*, Little, Brown and Company, New York

- [Gopalsamy et al. 1999]** C. Gopalsamy, S. Park, R. Rajamanickam, and S. Jayaraman (1999) "The Wearable Motherboard: The First Generation of Adaptive and Responsive Textile Structures (ARTS) for Medical Applications", *Virtual Reality*
- [Guy 1976]** W. Guy (1976) "Clinical Global Impressions", *ECDEU Assessment Manual for Psychopharmacology*, U.S. National Institutes of Health: Rockville, MD
- [Hamilton 1960]** M. Hamilton (1960) "A Rating Scale for Depression", *Journal of Neurology, Neurosurgery and Psychiatry*
- [Holter 1949]** N. Holter & J. Generelli (1949) "Remote Recording of Physiologic Data by Radio", *Rocky Mountain Med Journal*, pp. 747-751
- [Hsieh et al. 2003]** C. Hsieh, C. Mao, M. Young, T.Yeh, and S. Yeh (2003) "Respiratory Effect on the Pulse Spectrum", *Journal Medical Engineering Technology*
- [Hugdahl-A 1995]** K. Hugdahl (1995) *Psychophysiology: The Mind-Body Perspective*, Harvard University Press, Cambridge MA
- [Hugdahl-B 1995]** K. Hugdahl, (1995) *Electrodermal Activity, in Psychophysiology: The Mind-Body Perspective*, Harvard University Press, Cambridge, MA
- [Hugdahl-C 1995]** K. Hugdahl (1995) *Cardiovascular Psychophysiology, in Psychophysiology: The Mind-Body Perspective*", Harvard University Press, Cambridge MA
- [Icuiti]** Icuiti intelligent display solutions, <http://www.iisvr.com>
- [Intille 2002]** S. Intille, C. Kukla, and X. Ma (2002) "Eliciting User Preferences Using Image-Based Experience Sampling and Reflection", *Conference on Human Factors and Computing Systems*
- [Intille et al. 2003]** S. Intille, J. Rondoni, C. Kukla, I. Iacono, and L. Bao (2003) "A Context-Aware Experience Sampling Tool", *Conference on Human Factors and Computing*
- [Intille 2003]** S. Intille (2003) "Ubiquitous Computing Technology for Just-in-Time Motivation of Behavior Change", UbiHealth Workshop
- [Jaffe et al. 2001]** J. Jaffe, B. Beebe, S. Feldstein, C. Crown, and M. Jasnow (2001) "Rhythms of Dialogue in Early Infancy", *Monographs of the Society for Research in Child Development*, 66(2), No. 264
- [Jain et al. 2000]** A. Jain, R. Duin, and J. Mao (2000) "Statistical Pattern Recognition: A Review" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1): 4-37
- [Jakicic et al. 2004]** M. Jakicic, M. Marcus, K. Gallagher, C. Randall, E. Thomas, F. Goff, and R. Robertson (2004) "Evaluation of the SenseWear Pro Armband to Assess Energy Expenditure during Exercise", *Medicine & Science in Sports & Exercise*



- [Jensen 1996] F. Jensen (1996) *An Introduction to Bayesian Networks*, Springer-Verlag, London
- [Kawachi et al. 1995] I. Kawachi, D. Sparrow, P. Vokonas, and S. Weiss (1995) “Decreased Heart Rate Variability in Men with Phobic Anxiety”, *American Journal of Cardiology*, 75:882-885
- [Kendon et al. 1975] A. Kendon, R. Harris, and M. Key, Editors (1975) *Organization of Behaviour in Face to Face Interaction*, Mouton, The Hague, Netherlands
- [Kessler et al. 2003] R. Kessler et al. (2003) “The Epidemiology of Major Depressive Disorder: Results From the National Comorbidity Survey Replication (NCS-R)”, *Journal of the American Medical Association*, 289(23):3095-3105
- [Kirk et al. 2003] V. Kirk, S. Bohn, W. Flemons, and J. Remmers (2003) “Comparison of Home Oximetry Monitoring with Laboratory Polysomnography in Children”, *Chest*, 124:1702-1708
- [Klapper 2003] D. Klapper (2003) “Use of a Wearable Ambulatory Monitor in the Classification of Movement States in Parkinson’s Disease”, M.S. Thesis, Harvard-MIT Health Sciences and Technology Program
- [Kleiger et al. 1978] R. Kleiger and J. Miller (1978) “Decreased Heart Rate Variability and its Association with Increased Mortality after Acute Myocardial Infarction”, *Am J Cardiol*, 59:256-262
- [Kleiger et al. 1992] R. Kleiger, R. Stein, P. Bosner, and J. Rottman (1992) “Time Domain Measurements of Heart Rate Variability”, *Amb. Electrocardiol*, 10:487-498
- [Knight et al. 1994] G. Knight, L. Virdin, and M. Roosa (1994) “Socialization and Family Correlates of Mental Health Outcomes among Hispanic and Anglo American Children: Consideration of Cross-Ethnic Scalar Equivalence” *Child Dev.*, 65(1):212-224
- [Krause et al. 2003] Krause, A., Smailagic A., Siewiorek D., Farrington J., “Unsupervised, Dynamic Identification of Physiological and Activity Context in Wearable Computing” ISWC 2003
- [Kristal-Boneh et al. 1995] E. Kristal-Boneh, M. Raifel, P. Froom, and J. Ribak (1995) “Heart Rate Variability in Health and Disease”, *Scan. Journal Work Environment Health*, 21:85-95
- [Kumar & Pentland 2002] V. Kumar & A. Pentland (2002) “DiaBetNet: Learning and Predicting for Better Health and Compliance”, Diabetes Technology Society meeting, October, 2002
- [Kumar 2004] V. Kumar (2004) “The Design and Testing of a Personal Health System to Motivate Adherence to Intensive Diabetes Management”, M.S. Thesis, Harvard-MIT Health Sciences and Technology Program
- [Laerhoven et al. 2002] K. van Laerhoven, A. Schmidt, and H. Gellersen (2002) “Multi-sensor Context Aware Clothing”, *IEEE International Symposium on Wearable Computers*

[Lahmeyer & Bellur 1987] H. Lahmeyer & S. Bellur (1987) "Cardiac Regulation and Depression", *Journal of Psychiatric Research*

[Larson & Csikszentmihalyi 1983] R. Larson & M. Csikszentmihalyi (1983) "The Experience Sampling Method", In H. Reis, Editor, *Naturalistic Approaches to Studying Social Interaction: New Directions for Methodology of Social and Behavioral Science*, Jossey-Bass, San Francisco, CA

[Lisetti & Nasoz 2002] C. Lisetti & F. Nasoz (2002) "MAUI: A Multimodal Affective User Interface," *ACM Multimedia International Conference*

[Liu et al. 2003] K. Liu, P. Davis, and R. Picard (2003) "Wearable Relational Devices for Real-Time Annotation of Physiological Data", *The Science of Real-Time Data Capture: Self-Reports in Health Research Conference*

[Living Independently] Living Independently Corporation, <http://www.livingindependently.com>

[Lo & Repin 2002] A. Lo & D. Repin (2002) "The Psychophysiology of Real-Time Financial Risk Processing", *Journal of Cognitive Neuroscience*, 14:323-339

[Lo et al. 2005] A. Lo, D. Repin, and B. Steenbarger (2005) "Fear and Greed in Financial Markets: A Clinical Study of Day-Traders", *American Economic Review*

[Lobstein et al. 1983] D. Lobstein, B. Mosbacher, and A. Ismail (1983) "Depression as a Powerful Discriminator between Physically Active and Sedentary Middle-Aged Men", *J Psychosom Res.*, 27(1):69-76

[Lukowicz et al. 2002] P. Lukowicz, U. Anliker, J. Ward, G. Tröster, E. Hirt, and C. Neufelt (2002) "AMON: A Wearable Medical Computer for High Risk Patients", *IEEE International Symposium of Wearable Computers*

[Lynch et al. 2003] S. Lynch, J. Brown, and K. Harmsen (2003) "The Effect of Altering ADL Thresholds on Active Life Expectancy Estimates Among Older Persons", *Journal of Gerontology: Social Science*

[Lyyken et al. 1966] D. Lyyken, R. Rose, B. Luther, and M. Maley (1966) "Correcting Psychophysiological Measures for Individual Differences in Range", *Psychophysiological Bulletin*, 66:481-484

[Lyyken & Venables 1971] D. Lyyken & P. Venables (1971) "Direct Measurement of Skin Conductance: A Proposal for Standardization", *Psychophysiology*, 8(5):656-672

[Madan et al. 2004] A. Madan, R. Caneel, and Pentland (2004) "GroupMedia: Distributed Multimodal Interfaces", *International Conference on MultiModal Interfaces*

[Malan et al. 2004] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton (2004) "CodeBlue: An Ad Hoc Sensor Network Infrastructure for Emergency Medical Care", Workshop on Applications of Mobile Embedded Systems, *Mobisys*

**[Malliani et al. 1991]** A. Malliani, M. Pagani, F. Lombardi, R. Furlan, S. Guzzetti, and S. Cerutti (1991) "Spectral Analysis to Assess Increased Sympathetic Tone in Arterial Hypertension" *Hyperten*, 17(4):36-42

**[Marci 2002]** C. Marci (2002) "Psychophysiological Correlates of Empathy", Abstract, Harvard Myself Psychiatry Research Day, April, 2002

**[Markovitz et al. 1993]** J. Markovitz, K. Matthews, W. Kannel, and J. Cobb (1993) "Psychological Predictors of Hypertension in the Framingham Study: Is There Tension in Hypertension?", *Journal of the American Medical Association*, 270(20):2439-2494

**[Marmasse 2004]** N. Marmasse (2004) "Providing Lightweight Telepresence in Mobile Communication to Enhance Collaborative Living", Ph.D. Thesis, MIT Media Arts and Sciences

**[Micropaq]** Welch-Allyn's Micropaq wireless real-time monitor, <http://www.monitoring.welchallyn.com/products/wireless/micropaq.asp>

**[MiniMitter]** MiniMitter Corporation, <http://www.minimitter.com>

**[MIThril]** MIT MIThril Wearable Computing Project, <http://www.media.mit.edu/wearables/mithril>

**[MIThril Sensors]** MIThril sensors, <http://www.media.mit.edu/wearables/mithril/hardware/index.html>

**[Mundt 2002]** C. Mundt (2002) "Lifeguard- A Wearable Vitals Monitoring System", [http://lifeguard.stanford.edu/lifeguard\\_writeup\\_medium.pdf](http://lifeguard.stanford.edu/lifeguard_writeup_medium.pdf)

**[Murphy & Bird 1974]** R. Murphy & K. Bird (1974) "Telediagnosis: A New Community Health Resource; Observations on the Feasibility of Telediagnosis Based on 1000 Patient Transactions", *American Journal of Public Health*

**[Murphy 2001]** K. Murphy (2001) "The Bayes Net Toolkit for Matlab", <http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html>

**[Murray & Lopez 1996]** C. Murray & A. Lopez, Editors (1996) *The Global Burden of Disease and Injury Series, Volume 1: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*, Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank, Harvard University Press, Cambridge, MA

**[Nass & Brave 2004]** C. Nass & S. Brave (2004) *Voice Activated: How People Are Wired for Speech and How Computers Will Speak with Us*, MIT Press, Cambridge, MA

**[Nexense]** Nexense Doppler sensing technology, <http://www.nexense.com>

**[Nezlek et al. 1994]** J. Nezlek, M. Imbrie, and G. Shean, "Depression and Everyday Social Interaction", *J Pers Soc Psychol.*, 67(6):1101-1111

[Offerhaus 1980] R. Offerhaus (1980) "Heart Rate Variability in Psychiatry" in R. Kitney, O. Rompelman, Editors, *The Study of Heart Rate Variability*, Oxford University Press, Oxford

[Oppenheim & Schafer 1989] A. Oppenheim & R. Schafer (1989) *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ

[Oppenheimer & Cechetto 1990] S. Oppenheimer and D. Cechetto (1990) "Cardiac Chronotropic Organization of the Rat Insular Cortex", *Brain Research*

[Oppenheimer et al. 1992] S. Oppenheimer, et al. (1992) "Cardiovascular Effects of Human Insular Cortex Stimulation", *Neurology*

[Ori et al. 1992] Z. Ori, G. Monir, J. Weiss, X. Sayhouni, and D. Singer (1992) "Heart Rate Variability Frequency Domain Analysis", *Amb Electrocardio*, 10(3):499-537

[Oviatt & Cohen 2000] S. Oviatt & P. Cohen (2000) "Multimodal Interfaces That Process What Comes Naturally", *Communications of the ACM*

[Pascoe 1998] J. Pascoe (1998) "Adding Generic Contextual Capabilities to Wearable Computers", *IEEE International Symposium on Wearable Computers*

[Pentland 1998] A. Pentland (1998) "Wearable Intelligence," *Scientific American*

[Pentland 2003] A. Pentland (2003) "Social Dynamics: Signals and Behavior", *IEEE International Conference Developmental Learning*

[Pentland-A 2004] A. Pentland, "Healthwear: Medical Technology Becomes Wearable", *IEEE Computer*, May, 2004

[Pentland-B 2004] A. Pentland, "Social Dynamics: The Voice of Power and Influence, Submitted to: *International Conference of Multimodal Interfaces*, May, 2004

[Pentland et al. 2004] A. Pentland, J. Curhan, R. Khilnani, M. Martin, N. Eagle, R. Caneel, A. Madan (2004) "Toward a Negotiation Advisor", *ACM Symposium on User Interface and Software Technology*

[Picard 1997] R. Picard (1997) *Affective Computing*, MIT Press, Cambridge, MA

[Picard 2001] R. Picard (2001) "Affective Medicine: Technology with Emotional Intelligence", Chapter in *Future of Health Technology*, IOS Press

[Picard et al. 2001] R. Picard, E. Vyzas, and J. Healey (2001) "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State", *IEEE Transactions Pattern Analysis and Machine Intelligence*, 23:10

[Piette 2002] J. Piette (2002) "Enhancing Support via Interactive Technologies", *Current Diabetes Reports*, Vol. 2

[PlaceLab] house\_n PlaceLab, [http://architecture.mit.edu/house\\_n/web/placelab/PlaceLab.pdf](http://architecture.mit.edu/house_n/web/placelab/PlaceLab.pdf)

[Podlesny & Raskin 1977] J. Podlesny & D. Raskin (1977) "Physiological Measures and the Detection of Deception", *Psychol. Bull.*, 84(4):782-799

[Polar] Polar heart rate monitors, [www.polarusa.com](http://www.polarusa.com)

[Pozos & Danzl 2003] R. Pozos & D. Danzl (2003) "Human Physiological Responses to Cold Stress and Hypothermia," in *Medical Aspects of Harsh Environments Volume I*, Chapter 11, Borden Institute Walter Reed Army Medical Center

[Press & Rybicki 1989] W. Press and G. Rybicki, (1989) "Fast Algorithm for Spectral Analysis of Unevenly Sampled Data", *Astrophysical J.*, 338:277-280

[Qi et al. 2002] Y. Qi, T. Minka, and R. Picard (2002) "Bayesian Spectrum Estimation of Unevenly Sampled Nonstationary Data, *International Conference on Acoustics, Speech, and Signal Processing*

[Rabiner 1989] L. Rabiner (1989) "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition" *Proceedings of IEEE*, 7(2)

[Reiser 1978] S. Reiser (1978) *Medicine and the Reign of Technology*, Cambridge University Press, Cambridge, MA

[Remick et al. 2005] R. Remick, W. Linden, P. McClean, G. Iverson, and M Gaetz, (2005) "A New Potential Marker for Abnormal Cardiac Physiology in Depression.", *Journal of Behavioral Medicine*

[Riemann et al. 2001] D. Riemann, M Berger, and U. Voderholzer (2001) "Sleep and Depression – Results from Psychobiological Studies: An Overview", *Biol Psychol*, 57:67-103

[Rothschild et al. 1988] M. Rothschild, A. Rothschild, and M Pfeifer (1988) "Temporary Decrease in Cardiac Parasympathetic Tone after Acute Myocardial Infarction. *Am J Cardiol*, 62:637

[Rusting & Larsen 1998] C. Rusting & R. Larsen (1998) "Diurnal Patterns of Unpleasant Mood: Associations with Neuroticism, Depression, and Anxiety", *Journal of Personality*, 66(1):85

[Saul et al. 1988] P. Saul, Y. Arai, R. Berger, L. Lilly, W. Colucci, and R. Cohen (1998) "Assessment of Autonomic Regulation in Congestive Heartfailure by Heart Rate Spectral Analysis" *Am J Cardiol*, 61:1292-1299

[Saini et al. 1988] M. Saini, R. Kleiger, R. Carney, A. teVelde. and K. Freedland (1988) "Correlation of Heart Rate Variability with Clinical and Angiographic Variables and Late Mortality after Coronary Angiography", *Am J Cardiol*, 62:714-717

[Saint-Arnaud 1995] N. Saint-Arnaud (1995) "Classification of Sound Textures". M.S. Thesis, MIT Media Arts and Sciences

- [Schachter 2000] M. Schachter (2000) "The Diagnosis and Treatment of Hypothyroidism", Healthy.net Article, <http://www.healthy.net/scri/article.asp?PageType=Article&ID=528>
- [Schimmel 2001] M Schimmel (2001) "Emphasizing Difficulties in the Detection of Rhythms with Lomb-Scargle Periodograms", *Biol Rhythm. Res.*, 32(3):341-345
- [Sebastiani et al. 1999] P. Sebastiani, M. Ramoni, P. Cohen, J. Warwick, and J. Davis (1999) Discovering Dynamics Using Bayesian Clustering, *Lect Note Comput Sc*, 1642:199-210
- [Selye 1956] H. Selye (1956) *The Stress of Life*, Chapter 1-7. McGraw-Hill
- [Siuru 1997] B. Siuru (1997) "Applying Acoustic Monitoring to Medical Diagnostics Applications", *Sensor Magazine*, <http://www.sensormag.com/articles/0397/acoustic/main.shtml>
- [Slagle et al. 1989] J. Slagle, S. Finkelstein, L. Leung, and W. Warwick, "Monitor: An Expert System that Validates and Interprets Time-Dependent Partial Data Based on a Cystic Fibrosis Home Monitoring Program," *IEEE Transactions on Biomedical Engineering*, 1989
- [Sloan-A et al. 1994] R. Sloan, P. Shapiro, J. Bigger, E. Bagiella, R. Steinman, and J. Gorman (1994) Cardiac Autonomic Control and Hostility in Healthy Subjects", *Am J Cardiol*, 74:298-300
- [Sloan-B et al. 1994] R. Sloan, P. Shapiro, E. Bagiella, S. Boni, M. Paik, R. Steinman, and J. Gorman (1994) "Effect of Mental Stress Throughout the Day on Cardiac Autonomic Control", *Bio Psychol* 37:89-99
- [Sobin & Sackeim 1997] C. Sobin, H. Sackeim (1997) "Psychomotor Symptoms of Depression", *Am J Psychiatry*, 154(1):4-17
- [Spot] Carnegie Mellon Spot Project, <http://www.wearablegroup.org/hardware/spot>
- [Stampfer 1998] H. Stampfer: "Relationship between Psychiatric Illness and Circadian Pattern of Heart Rate" *Journal of Psychiatry*, 32:187-198
- [Stampfer et al. 2002] H. Stampfer, G. Iverson (2002) "Reliability of Circadian Heart Rate Patterns in Psychiatry" (73):3
- [Starner et al. 1998] T. Starner, B. Schiele, and A. Pentland (1998) "Visual Context Awareness in Wearable Computing", IEEE International Symposium on Wearable Computers
- [Starner 2001] T. Starner (2001) "The Challenges of Wearable Computing: Part 2," IEEE Micro
- [Steenbarger 2002] B. Steenbarger (2002) *The Psychology of Trading: Tools and Techniques for Minding the Markets*, John Wiley & Sons, Hoboken, NJ
- [Stock et al. 2004] J. Stock, N. Taylor, M. Tipton, and J. Greenleaf (2004) "Human Physiological Responses to Cold Exposure" *Aviation, Space, and Environmental Medicine*
- [Strang 1988] G. Strang (1988) *Linear Algebra and its Applications*, 3<sup>rd</sup> Edition, Saunders, Philadelphia, PA

- [Sung 2002] M. Sung (2002) VitaMon: Critical Health Monitoring, Technical Report, MIT Media Laboratory
- [Sung 2003] M. Sung (2003) "Shivering Motion/Hypothermia Classification for Wearable Soldier Health Monitoring Systems," *IEEE International Symposium of Wearable Computers*
- [Sung & Gips 2003] M. Sung & J. Gips (2003) "Context-Aware Accelerometer-Based Head Nodding/Shaking Classifier", Technical Report, MIT Media Laboratory
- [Sung 2004] M. Sung (2004) "Non-Invasive Activity Classification Using Single Accelerometer", Technical Report, MIT Media Laboratory
- [Sung & Pentland 2004] M. Sung & A. Pentland (2004) "MITHril LiveNet: Health and Lifestyle Networking", Workshop on Applications of Mobile Embedded Systems, *MobiSys*
- [Sung et al. 2004] M. Sung, R. DeVaul, S. Jimenez, J. Gips, and A. Pentland (2004) "A Shiver Motion and Core Body Temperature Classification for Wearable Soldier Health Monitoring Systems" *IEEE International Symposium of Wearable Computers*
- [Sung & Pentland 2005] M. Sung & A. Pentland (2005) "Non-Invasive Physiologic Sensing for Human-Aware Interfaces", HCI International, *International Conference on Universal Access in Human-Computer Interaction*
- [Sung-A et al. 2005] M. Sung, J. Gips, N. Eagle, A. Madan, R. Caneel, R. DeVaul, J. Bonsen, and A. Pentland (2005) "MIT.EDU: M-Learning Applications for Classroom Settings", *Journal of Computer Assisted Learning*, 21:3
- [Sung-B et al. 2005] M. Sung, C. Marci, and A. Pentland (2005) "Wearable Feedback Systems for Rehabilitation", *Journal of NeuroEngineering and Rehabilitation*, 2:17
- [Szekely 2000] M. Szekely (2000) "The Vagus Nerve in Thermoregulation and Energy Metabolism", *Auton Neurosci*, 85(1-3):26-38
- [Szolovits et al. 1994] P. Szolovits, J. Doyle, W. Long, I. Kohane, and S. Pauker (1994) "Guardian Angel: Patient-Centered Health Information Systems", Technical Report No.: MIT/LCS/TR-604, MIT Laboratory for Computer Science
- [Taillard et al. 1990] J. Taillard, et al. (1990) "Heart Rate Circadian Rhythm as a Biological Marker of Desynchronization in Major Depression: A Methodological and Preliminary Report", *Chronobiology International*
- [Taillard et al. 1993] J. Taillard, et al. (1993) "Sleep and heart Rate Circadian Rhythm in Depression", *Chronobiology International*
- [Tapia et al. 2004] E. Tapia, N. Marmasse, S. Intille, and K. Larson (2004) "MITes: Wireless Portable Sensors for Studying Behavior", Technical Report, House\_n Group, MIT Media Laboratory

**[Tavormina 1999]** C. Tavormina (1999) “Embracing the Eden Alternative in Long-Term Care Environments”, *Geriatr Nurs.*, (3):158-161

**[TFESC 1996]** Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) “Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical Use”, *Circulation*, 92:1043-1065

**[Thorell-A et al. 1987]** L. Thorell, B. Kjellman, and G. D'Elia (1987) “Electrodermal Activity in Relation to Diagnostic Subgroups and Symptoms of Depressive Patients”, *Acta Psychiatrica Scandinavica*

**[Thorell-B et al. 1987]** L. Thorell, B. Kjellman, and G. D'Elia (1987) “Electrodermal Activity in Antidepressant Medicated and Unmedicated Depressive Patients and in Matched Healthy Subjects”, *Acta Psychiatrica Scandinavica*

**[Thorell & d'Elia 1988]** L. Thorell & G. d'Elia (1988) “Electrodermal Activity in Depressive Patients in Remission and in Matched Healthy Subjects”, *Acta Psychiatrica Scandinavica*

**[Tipping & Bishop 1997]** M. Tipping & C. Bishop (1997) “Probabilistic Principal Component Analysis”, Technical Report NCRG/97/010, Neural Computing Research Group, Aston University

**[Tsai et al. 2003]** J. Tsai, et al. (2003) “The Effects of Depression on the Emotional Responses of Spanish-speaking Latinas”, *Cultural Diversity & Ethnic Minority Psychology*

**[Tsujimoto et al. 1990]** T. Tsujimoto, N. Yamada, K. Shimoda, K. Hanada, S. Takahashi (1990) “Circadian Rhythms in Depression. Part II: Circadian Rhythms in Inpatients with Various Mental Disorders”, *J Affect Disord.*,18(3):199-210

**[Turk & Pentland 1998]** M. Turk and A. Pentland (1991) “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, 3(1)

**[Uber-Badge]** Uber-Badge Project, <http://www.media.MIThrl/resenv/badge/>

**[URCFH]** University of Rochester Center for Future Health Smart Home, [http://www.futurehealth.rochester.edu/smart\\_home](http://www.futurehealth.rochester.edu/smart_home)

**[Volkers et al. 2003]** A. Volkers, et al (2003) “Motor Activity and Autonomic Cardiac Functioning in Major Depressive Disorder”, *Journal of Affective Disorders*

**[Vyzas & Picard 1998]** E. Vyzas & R. Picard (1998) “Affective Pattern Classification,” *AAAI Symp., Emotional and Intelligent: The Tangled Knot of Cognition*

**[Vyzas & Picard 1999]** E. Vyzas & R. Picard (1999) "Offline and Online Recognition of Emotion Expression from Physiological Data", Workshop on Emotion-Based Agent Architectures, *International Conference on Autonomous Agents*, pp. 135-142



[Ward et al. 1983] N. Ward, H. Doerr, and M. Storrie (1983) "Skin Conductance: A Potentially Sensitive Test for Depression", *Psychiatry Research*

[Warner et al. 1987] R. Warner, et al. (1987) "Rhythmic Organization of Social Interaction and Observer Ratings of Positive Affect and Involvement", *Journal of Nonverbal Behavior*

[Warner 1992] R. Warner (1992) "Dimensions of Social Interaction Tempo: A Factor Analytic Study of Time and Frequency Domain Indexes Of Interaction Structure", *Journal of Psycholinguistic Research*

[Watson et al. 1995] D. Watson, K. Weber, J. Assenheimer, L. Clark, M. Strauss, and R. McCormick (1995) "Testing a Tripartite Model: I. Evaluating the Convergent and Discriminant Validity of Anxiety and Depression Symptom Scales", *J Abnorm Psychol.*, 104(1):3-14

[Watt & VanLear 1996] J. Watt & C. VanLear (1996) *Dynamic Patterns in Communication*, Sage Publications, Thousand Oaks, CA

[Weaver 2003] J. Weaver (2003) "A Wearable Health Monitor to Aid Parkinson Disease Treatment," M.S. Thesis, MIT Media Arts and Sciences

[Wertheimer & Santella 2001] A. Wertheimer & T. Santella (2001) "Medication Compliance Research: Still So Far to Go", *Journal of Applied Research in Clinical and Experimental Therapeutics*

[Wiener et al. 1990] J. Wiener, R. Hanley, R. Clark, and J. Van Nostrand (1990) "Measuring the Activities of Daily Living: Comparisons Across National Surveys", *Journal of Gerontology: Social Sciences*

[Wilson 1999] C. Wilson (1999) Hospitals of the Future: The Impact of Medical Technologies on the Future of Hospitals, *British Medical Journal*, 319:1287

[Wren & Pentland 1999] C. Wren & A. Pentland, "Understanding Purposeful Human Motion", *IEEE Face and Gesture Recognition*

[Wysocki et al. 2003] T. Wysocki, M. Harris, K. Wilkinson, M. Sadler, N. Mauras, and N. White (2003) "Self-Management Competence as a Predictor of Outcomes of Intensive Therapy or Usual Care in Youth with Type 1 Diabetes", *Diabetes Care*

[Xu & Jordan 1996] L. Xu. & M. Jordan (1996) "On Convergence Properties of the EM Algorithm for Gaussian Mixtures", *Journal of Neural Computation*, 8(1):129-151

[Yeragani et al. 1991] V. Yeragani, R. Pohl, R. Balon, C. Ramesh, D. Glitz, I. Jung and P. Sherwood (1991) "Heart Rate Variability in Patients with Major Depression", *Psychiat Res*, 37:35-46

[Yeragani et al. 1990] V. Yeragani, R. Balon, R. Pohl, C. Ramesh, D. Glitz, P. Weinberg, B. Merlos (1990) "Decreased R-R Variance in Panic Disorder Patients", *Acta Psychiatr Scand*, 81:554-559

**[Yeragani et al. 1993]** V. Yeragani, R. Pohl, R. Berger, R. Balon, C. Ramesh, D. Glitz, K. Srinivasan, and P. Weinberg (1993) "Decreased Heart Rate Variability in Panic Disorder Patients: A Study of Power-Spectral Analysis of Heart Rate", *Psychiatry Research*, 46:89-103

**[Zung et al. 1990]** W. Zung, K. Magruder-Habib, R. Velez, W. Alling (1990) "The Comorbidity of Anxiety and Depression in General Medical Patients: A Longitudinal Study", *J Clin Psychiatry*, 51 Suppl:77-80; Discussion 81