

# **Creating Cohesive Video with the Narrative-Informed use of Ubiquitous Wearable and Imaging Sensor Networks**

by

**Mathew Laibowitz**

S.M., Massachusetts Institute of Technology (2004)

B.S., Columbia University (1997)

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning,  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 2010

© Massachusetts Institute of Technology 2010. All rights reserved.

Author \_\_\_\_\_  
Program in Media Arts and Sciences  
January 26, 2010

Certified by \_\_\_\_\_  
Joseph A. Paradiso  
Professor of Media Arts and Sciences  
MIT Media Lab  
Thesis Supervisor

Accepted by \_\_\_\_\_  
Deb K. Roy  
Chair  
Departmental Committee on Graduate Students



# **Creating Cohesive Video with the Narrative-Informed use of Ubiquitous Wearable and Imaging Sensor Networks**

by

**Mathew Laibowitz**

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning,  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in Media Arts and Sciences

## **Abstract**

In today's digital era, elements of anyone's life can be captured, by themselves or others, and be instantly broadcast. With little or no regulation on the proliferation of camera technology and the increasing use of video for social communication, entertainment, and education, we have undoubtedly entered the age of ubiquitous media. A world permeated by connected video devices promises a more democratized approach to mass-media culture, enabling anyone to create and distribute personalized content. While these advancements present a plethora of possibilities, they are not without potential negative effects, particularly with regard to privacy, ownership, and the general decrease in quality associated with minimal barriers to entry.

This dissertation presents a first-of-its-kind research platform designed to investigate the world of ubiquitous video devices in order to confront inherent problems and create new media applications. This system takes a novel approach to the creation of user-generated, documentary video by augmenting a network of video cameras integrated into the environment with on-body sensing. The distributed video camera network can record the entire life of anyone within its coverage range and it will be shown that it, almost instantly, records more audio and video than can be viewed without prohibitive human resource cost. This drives the need to develop a mechanism to automatically understand the raw audio-visual information in order to create a cohesive video output that is understandable, informative, and/or enjoyable to its human audience.

We address this need with the SPINNER system. As humans, we are inherently able to transform disconnected occurrences and ideas into cohesive narratives as a method to understand, remember, and communicate meaning. The design of the SPINNER application and ubiquitous sensor platform is informed by research into narratology, in other words how stories are created from fragmented events. The SPINNER system maps low level sensor data from the wearable sensors to higher level social signal and body language information. This information is used to label the raw video data. The SPINNER system can then build a cohesive narrative by stitching together the appropriately labeled video segments.

The results from three test runs are shown, each resulting in one or more automatically edited video piece. The creation of these videos is evaluated through review by their intended audience and by comparing the system to a human trying to perform similar actions. In addition, the mapping of the wearable sensor data to meaningful information is evaluated by comparing the calculated results to those from human observation of the actual video.

Thesis Supervisor: Joseph A. Paradiso  
Title: Professor, Program in Media Arts and Sciences





# **Creating Cohesive Video with the Narrative-Informed use of Ubiquitous Wearable and Imaging Sensor Networks**

by

**Mathew Laibowitz**

The following people served as readers for this thesis:

Thesis Reader \_\_\_\_\_

Alex “Sandy” Pentland  
Professor of Media Arts and Sciences  
MIT Media Lab

Thesis Reader \_\_\_\_\_

David Rakoff  
Essayist, Author, Actor  
Bestselling Author of “Fraud” and “Don’t Get Too Comfortable”

Thesis Reader (IN MEMORIAM) \_\_\_\_\_

Blake Snyder  
Screenwriter, Author of “Save The Cat”  
Screenwriting Lecturer, UCLA



## Acknowledgments

---

To my advisor Joseph Paradiso for the seven years of Italian Prog, and all the inspiration, advice, and help without which I would never have done this

To my illustrious committee Alex “Sandy” Pentland and David Rakoff, thanks for badges a-plenty and the amazing duct tape wallet, it was an honor and pleasure to have you involved with my work and my life

To all the professor I have had here, Glorianna Davenport, Tod Machover, Hiroshi Ishii, and Allan McCollum, I have used and will continue to use the lessons that you taught me everyday

To the Things That Think Consortium and the Media Lab sponsors who have funded me directly or indirectly

To the Media Lab community for your understanding of the drilling, screwing, and destruction

To Facilities and NECSYS, without whose support this would not have been possible

To Nokia for the funding support

To TI for the development tools

To Michael Chang and the rest of the Empower Team, thanks for the endless support and advice

Thanks to Lisa Lieberman for keeping the ship afloat, and always being ready to help me out and get things done

To those that have come before me...

Thanks to Ari Benbesat for keeping things light with non-light humor, my kind of humor

Thanks to Josh Lifton for locking up everything that smells in a box, midnight madness, European adventures, and the list keeps going

Thanks to Mark Feldmeier for saving me when all was not well with me, and for song poem day

Thanks to Bo Morgan for writing SPINNERD and the heroic 7-11 run on January 25th

Thanks to Nan-wei Gong for all the support with this project, for the frog, and for the amazing effort that I can always count on

Thanks to Mike Lapinski for running the deployment and Matt Aldrich for three tours, and Alex Reben for all the SPINNER help

And to the Responsive Environments Group, past and present, Gershon Dublon, Laurel Pardue, Clemens Setzger, Hong Ma, Behram Mistree, Ryan Aylward, Gerardo Berota, David Merrill, Manas Mittal, Jason LaPenta, Xmal, Brox, Secret Asian Dan, Stacey Morris

Thanks to Peggy for cleaning up after me. The peanuts were not just Josh's

Thanks to Suelin Chen for spending 30 hours watching random video footage for a single point of evaluation, and for your devotion and love for the past four and a half years

Thanks to Alison Hammer for the graphic design help, the sandwiches, and the awesome roommating

To Mario Bava, Lucio Fulci, Claudio Simonetti, and Dario Argento for the inspiration

To my dearest friends for all the support and understanding, I am so excited to finally have the time to visit, Doron, Tim, Anna Rae, Suneil, Nick, Cathy, Nicole, Lisa, Abby, Corey, Arin, Jay, Dan, Tamara, Kamallesh, Vinit, Nancy, Diana, Andy, Alyssa, Bryan, Josh H, Shauna, Anthony, Rebecca, Adam

To my family, thanks for 34 years of immeasurable support, pride without question, for whom this accomplishment is dedicated, without whom I would not have been able, nor would I have valued, the completion of this challenge, Mom + Dad, Ken + Danielle, Nora + David, Ben, Hannah, Rebecca, Juliana, and Rachel. And to my grandmother Giuliana Ravenna Goetzl, who passed away during the course of this work, so much of this comes from you

R.I.P. Blake Snyder, you will be sorely missed, this story is not over



# CONTENTS

.....

The Spoiler follows you  
In Spoiler's time  
from place to place  
Old man's eyes  
In a young boy's face  
Walking in  
Another's skin  
Secrets inside  
The Spoiler  
The Spoiler's catch  
Boys who exist  
In the stories of shadows

- Coil [1]

The *Contents* chapter contains the spoilers, in other words the Table of Contents and List of Figures.

## TABLE OF CONTENTS

Contents .....	9
Table of Contents .....	10
List of Figures .....	15
List of Tables.....	21
 Chapter 1 - Antenarrative .....	 23
1.1 The Age of Ubiquitous Media.....	24
1.1.1 Ubiquitous Media Consumption .....	24
1.1.2 Ubiquitous Media Production .....	28
1.1.3 Additional Attributes of the Ubiquitous Age .....	30
1.1.4 Research Requirements .....	31
1.2 Distributed Media System.....	32
1.2.1 Obstacles to the Creation of this type of Platform .....	34
1.2.2 SPINNER Media Platform .....	35
1.3 Wearable Sensors .....	36
1.4 Identifying Meaningful Content.....	38
1.5 Introduction to Applications .....	41
1.5.1 Labeling and Browsing .....	41
1.5.2 SPINNER Application .....	41
1.5.3 Searching the Future.....	42
1.5.4 Sensor-Augmented Photography.....	43
1.5.5 Visualization.....	43
1.5.6 Situated Interactive Applications .....	44
 Chapter 2 - Prologue .....	 45

2.1	Related Work .....	46
2.1.1	Video Sensor Networks.....	46
2.1.2	Human-centric Sensing .....	47
2.1.3	Meaningful Content vs. Discordant Events.....	48
 Chapter 3 - System Design .....		51
3.1	Overall System Architecture.....	52
3.2	Distributed Media System.....	53
3.2.1	Modular Component #1 - The Red Board.....	55
3.2.1.1	Red Board Sensors.....	55
3.2.1.2	Red Board Audio System .....	57
3.2.1.3	Red Board Processor.....	57
3.2.1.4	Red Board ZMat Transceiver .....	57
3.2.1.5	Red Board Infrared System .....	59
3.2.1.6	Red Board Wired Interfaces .....	59
3.2.1.7	Red Board Firmware.....	61
3.2.2	Modular Component #2 - The Green Board .....	62
3.2.2.1	The Green Board - Hardware Design .....	63
3.2.2.2	The Green Board - Operating System .....	66
3.2.2.3	The Green Board - DSP/Video Code.....	67
3.2.2.4	Application Services.....	69
3.2.3	Modular Component #3 - The Hat .....	72
3.2.4	Modular Component #4 - The Power Supply .....	74
3.2.5	Complete Portal Assembly.....	75
3.3	Wearable Sensor Devices.....	81
3.3.1	The $\mu$ Badge .....	81
3.3.1.1	microBadge Sensor Suite.....	85
3.3.1.2	microBadge Audio System.....	86
3.3.1.3	microBadge ZMat Wireless Transceiver .....	87
3.3.1.4	microBadge Additional Features .....	88
3.3.2	Wrist Sensor Unit.....	89
3.3.2.1	Wrist Unit Sensor Suite .....	93
3.3.3	Battery Life .....	94
3.4	Backend Services .....	95

3.4.1	Relational Database .....	96
3.4.2	SPINNER Daemon .....	100
3.5	Platform Applications .....	102
3.5.1	Privacy Study .....	102
3.5.2	Awareness Application .....	103
3.5.3	Visualization and Mobile Browsing.....	105
3.5.4	Cross Reality .....	106
3.5.5	Cloud of Images .....	108

## Chapter 4 - SPINNER Video Application .....111

4.1	Introducing SPINNER Video.....	112
4.2	Narrative-Informed Design .....	113
4.3	Video Segmentation .....	119
4.3.1	Location Engine .....	120
4.3.2	Social Situation Detection.....	124
4.3.3	Labeling of Clips .....	126
4.3.4	Evaluation of Clip Detection.....	127
4.4	Input Parameters.....	128
4.4.1	Story Arc Parameters .....	128
4.4.2	Pacing Parameters .....	129
4.4.3	Music mappings .....	130
4.4.4	Scripting Specific Points .....	130
4.4.5	Story Energy Mapping .....	131
4.5	Final Video Output.....	131

## Chapter 5 - Sensor Data Processing .....133

5.1	Human Signal Processing .....	134
5.1.1	Developing the Mapping.....	135
5.1.2	Summary of Mapped Parameters .....	138
5.2	Mapping Details and Test Results.....	142



5.2.1	Social Interaction Dynamics .....	142
5.2.2	Activity.....	147
5.2.3	Social Signals .....	153
5.2.3.1	Speaker and Listener Intensity.....	154
5.2.4	Speech Tonal Analysis .....	166
5.2.5	Summary Diagram of Sensor Data Processing .....	168
5.3	Evaluation of Mapping Accuracy .....	169
Chapter 6 - SPINNER Video Results .....		175
6.1	Application Setup.....	176
6.2	Video #1 - Montage .....	176
6.3	Video #2 - Journey .....	179
6.4	Video #3 - Story Arc.....	180
6.5	Evaluation of Video Results .....	183
Chapter 7 - Epilogue .....		189
7.1	Contributions.....	190
7.2	Future Work .....	192
7.3	Final Thoughts .....	194
Appendices .....		195
A.1	Schematics.....	196
A.1.1	Red Board - CPU.....	196
A.1.2	Red Board - Ethernet.....	197
A.1.3	Red Board - Peripherals .....	198
A.1.4	Red Board - Power .....	199
A.1.5	Red Board - Radio.....	200
A.1.6	Hat - LEDs and Controller .....	201
A.1.7	Hat - Motor Controller .....	202

A.1.8	Hat - Microcontroller.....	203
A.1.9	Camera Cable and Buffers .....	204
A.1.10	Speaker Amplifier .....	205
A.1.11	Power Supply .....	206
A.1.12	Green Board .....	207
A.1.13	Wrist - Galvanic Skin Response Sensor.....	207
A.1.14	Wrist - Processor .....	208
A.1.15	Wrist - OLED Display.....	209
A.1.16	Wrist - Inertial Measurement Unit .....	210
A.1.17	Wrist - Peripherals.....	211
A.1.18	Wrist - Power Supply, Battery Charge Controller .....	212
A.1.19	Wrist - Radio .....	213
A.1.20	µBadge - Processor.....	214
A.1.21	µBadge - OLED Display .....	215
A.1.22	µBadge - Inertial Measurement Unit.....	216
A.1.23	µBadge - Peripherals .....	217
A.1.24	µBadge - Power Supply, Battery Charge Controller.....	218
A.1.25	µBadge - Radio.....	219
A.2	Source Code .....	220

References .....	221
------------------	-----

## LIST OF FIGURES

FIGURE 1-1 Current and predicted growth of UGC creation .....	26
FIGURE 3-2 Multi-tiered system.....	52
FIGURE 3-3 Sensate Ubiquitous Media Portal .....	53
FIGURE 3-4 Ubiquitous Media Portal Block Diagram.....	54
FIGURE 3-5 Component diagram of situated sensor node, a.k.a. The Red Board ....	56
FIGURE 3-6 The ZigBee mesh network topology containing a single coordinator, a multitude of routers, and mobile end-devices. ....	58
FIGURE 3-7 The Red Board with flying wires, ready to be integrated into a Ubiquitous Sensor Portal .....	60
FIGURE 3-8 Davinci Block Diagram [77] .....	63
FIGURE 3-9 Front side of the Green Board .....	64
FIGURE 3-10 Rear Side of the Green Board.....	65
FIGURE 3-11 The Camera Module on its custom cable .....	66
FIGURE 3-12 Davinci Linux kernel source GIT tree.....	67
FIGURE 3-13 Davinci Software Components Topology [77] .....	68
FIGURE 3-14 Hat module with faceplate.....	72
FIGURE 3-15 Front of Hat board with ganged high-brightness LEDs .....	73
FIGURE 3-16 Back of Hat board with electronics .....	73
FIGURE 3-17 Power Supply Board Diagram.....	74
FIGURE 3-18 Wall Mount .....	76
FIGURE 3-19 Green Board attached to side pieces.....	77
FIGURE 3-20 Front plate with attached speaker, LCD, and panning camera.....	78
FIGURE 3-21 Front plate with Red Board ready to assemble.....	78
FIGURE 3-22 Side view of complete mounted assembly with armor.....	79
FIGURE 3-23 Portal in action, illuminating subject for video capture .....	80
FIGURE 3-24 Completed Ubiquitous Media Portal.....	80
FIGURE 3-25 SPINNER Wearable sensors .....	81
FIGURE 3-26 μBadge pinned to participant .....	82

FIGURE 3-27	Diagram of the microBadge (front) .....	83
FIGURE 3-28	Diagram of the microBadge (back) .....	84
FIGURE 3-29	μΒαδγϵ in its case, close-up .....	87
FIGURE 3-30	The wrist sensor unit has a speaker, microphone, display, buttons, and wireless communication, making it quite similar to the fictional Dick Tracy wrist communication device [79] .....	89
FIGURE 3-31	Wrist Sensor Unit .....	89
FIGURE 3-32	CAD rendering of the case top .....	90
FIGURE 3-33	CAD rendering of case bottom prior to printing .....	90
FIGURE 3-34	Diagram of wrist sensor unit (front) .....	91
FIGURE 3-35	Diagram of wrist sensor unit (back) .....	92
FIGURE 3-36	Wrist unit's GSR probes on arm side of device .....	93
FIGURE 3-37	SPINNER Database Schema Diagram .....	97
FIGURE 3-38	SPINNER Database MySQL sensor data query view .....	98
FIGURE 3-39	SPINNER Data/Video Viewer .....	99
FIGURE 3-40	Spinnerd status page .....	100
FIGURE 3-41	Spinnerd node report .....	101
FIGURE 3-42	Nan-wei Gong's Privacy System and data collected from user tests [28] 103	
FIGURE 3-43	On air indication .....	104
FIGURE 3-44	Portal running Awareness Application.....	104
FIGURE 3-45	Portal displaying a movie .....	104
FIGURE 3-46	Mobile Phone SPINNER viewer .....	105
FIGURE 3-47	Tangible Ubiquitous Media Explorer .....	105
FIGURE 3-48	UbiCorder .....	106
FIGURE 3-49	Portal displaying the virtual world view.....	107
FIGURE 3-50	Drew Harry's 3D SPINNER-based video library application.....	107
FIGURE 3-51	Virtual Media Lab showing a portal into the real Media Lab (center right) 108	
FIGURE 3-52	Cloud of Images showing RSS and the last captured image .....	109
FIGURE 4-53	Stories can be metaphorically seen as the first derivative of reality, which describes only the changes in reality, but in such a way that the overall experi-	

ence can be communicated fully and efficiently .....	113
FIGURE 4-54 Vonnegut's Cinderella story structure. Specific story moments are described where a major change in audience experience happens. Note that, although they are specific to Cinderella, they are very easily adapted to describe any event in any story.....	115
FIGURE 4-55 Vonnegut general disaster story structure .....	116
FIGURE 4-56 Standard 3-act story arc .....	116
FIGURE 4-57 Example of story arc and character arcs.....	117
FIGURE 4-58 Kurt Vonnegut's Real-Life Story Model, illustrating the sparseness of actual change .....	118
FIGURE 4-59 Map of camera locations during test run .....	120
FIGURE 4-60 Output of location system during test run .....	121
FIGURE 4-61 Sample wearable node trajectory visualized from collected location data .....	122
FIGURE 4-62 Diagram of the IR social interaction detection.....	124
FIGURE 5-63 Non-Verbal Communication .....	137
FIGURE 5-64 SPINNER Wearable Devices .....	138
FIGURE 5-65 Audio waveforms from the badges of participant 6(top) and participant 7(bottom), displaying a conversation from the 1 hour test .....	145
FIGURE 5-66 Audio energy envelopes illustrating the conversational dynamics between participants 6 (top) and 7 (bottom) .....	146
FIGURE 5-67 Badge Gyro Axes .....	148
FIGURE 5-68 Badge Accelerometer Axes .....	148
FIGURE 5-69 Example activity plot from the Body Accelerometer Magnitude .....	149
FIGURE 5-70 Example activity plot from the Wrist Accelerometer Magnitude .....	150
FIGURE 5-71 Overall Activity Envelope.....	152
FIGURE 5-72 GSR Response of Participant 3 .....	152
FIGURE 5-73 Hand activity during listening highlighted for two participants in a conversation, audio envelope and hand motion shown, data taken from actual natural test situation. Shaded areas are time periods automatically selected when hand motion is high and speech envelope is low. These areas are considered as times of low attentiveness due to hand motion while listening.....	155

<b>FIGURE 5-74</b> Posture angle of two conversation participants, one shows a smooth bobbing with a generally steady trend, the other shows significant bobbing with an emphatic finish to the conversation .....	156
<b>FIGURE 5-75</b> Compass angles of speaker and listener, and their relative angle, these two participants show a very low deviation angle indicating attentiveness. Towards the end of the conversation, the deviation angle shows motion, indicating a change in conversational state, most likely from some sort of goodbye gesture.....	157
<b>FIGURE 5-76</b> Cross-correlation between two people in a conversation, known to be engaged in the conversation and mirroring to some extent.....	158
<b>FIGURE 5-77</b> Mirroring score of above cross-correlation calculated by averaging the columns.....	159
<b>FIGURE 5-78</b> Cross-correlation and mirroring score between two people in a conversation, known to be disengaged and not mirroring until the end.....	159
<b>FIGURE 5-79</b> The jitter is calculated from the body motion intensity while listening with the mirroring score subtracted. So the higher the mirroring the lower the jitter. Due to the large window required to calculate mirroring, there is a delay in the mirroring plot. Both the jitter value and the mirroring value while listening are used in the listener energy calculation.....	160
<b>FIGURE 5-80</b> Hand activity while speaking is highlighted. The amount of hand motion delivered while speaking is a good indication of the amount of emphasis the speaker is putting behind his/her speech. The green areas on the plot are automatically created to show areas of emphatic gesture. The level of hand motion intensity while speaking is the main component of the speaking energy parameter .....	161
<b>FIGURE 5-81</b> Participant 1's Speech Intensity and Listening Intensity/Attentiveness, the value is set to zero for speaking intensity while listening and set to zero for listening intensity while speaking.....	162
<b>FIGURE 5-82</b> Participant 2's Speech Intensity and Listening Intensity.....	163
<b>FIGURE 5-83</b> The conversation energy of Participant 1 is calculated by adding the zero-aligned speaking and listening energy values .....	164
<b>FIGURE 5-84</b> The relationship score over the course of one conversation between Participant 1 and Participant 2. You can see the aforementioned spike at 90 seconds, but in general the conversation goes downhill from there and dwindles	

into a fairly indifferent interaction .....	165
<b>FIGURE 5-85</b> Example of the most common relationship curve, quite flat .....	166
<b>FIGURE 5-86</b> Vocal Stress Analysis using Badges during a Conversation. The participant whose stress analysis results are in the top plot was asked to try and stress the other participant out. When you remove the bias from Participant 1's data, you can see significantly higher peaks in Participant 2's stress analysis results	167
<b>FIGURE 5-87</b> Block Diagram Summarizing the Processing from Raw Input to Outputs Available for Mapping to Story Energy and/or Labeling Video .....	168
<b>FIGURE 5-88</b> Charted Survey Results .....	173
<b>FIGURE 6-89</b> Montage Video Trajectory .....	177
<b>FIGURE 6-90</b> Montage Video Still .....	178
<b>FIGURE 6-91</b> Journey Video Still .....	180
<b>FIGURE 6-92</b> Story Arc Video Trajectory .....	181
<b>FIGURE 6-93</b> Story Arc Video Still, showing the main character at the height of searching. This is along the steep upward section of the story trajectory, just before he locates the social situation that he is searching for .....	182
<b>FIGURE 6-94</b> Story Arc Video Still, showing the height of social activity at the peak of the trajectory curve .....	183
<b>FIGURE 7-95</b> Keep on SPINNING! .....	194





# LIST OF TABLES



TABLE 4-1 Percentage of time spent in front of camera during test ..... 123

TABLE 4-2 Social activity during 1-hour test..... 125

TABLE 4-3 Example EDL ..... 132

TABLE 5-4 Sensor Mapping Questionnaire (page 1) ..... 169

TABLE 5-5 Sensor Mapping Questionnaire (page 1 cont)..... 170

TABLE 5-6 Sensor Mapping Questionnaire (page 2) ..... 171

TABLE 5-7 Results from sensor data mapping evaluation ..... 172

TABLE 5-8 Results from sensor data mapping evaluation (cont)..... 172

TABLE 6-9 SPINNER Review Questions ..... 184

TABLE 6-10 SPINNER Review Questions (cont) ..... 185

TABLE 6-11 SPINNER Review Survey Results..... 186



# ANTENARRATIVE

A decorative dotted line consisting of a horizontal row of dots followed by a vertical column of dots, forming an L-shape.

**S**torytelling reveals meaning without committing the error of defining it.

- Hannah Arendt [2]

The *Antenarrative* chapter, commonly called the introduction, summarizes and presents the entire course of this dissertation. The research program described herein can itself be seen as a sequence of events with an overall goal and cohesive meaning, in other words, a narrative. The *Antenarrative* is therefore the collective ideas and thoughts from which the structured story ultimately unfolds.

## 1.1 THE AGE OF UBIQUITOUS MEDIA

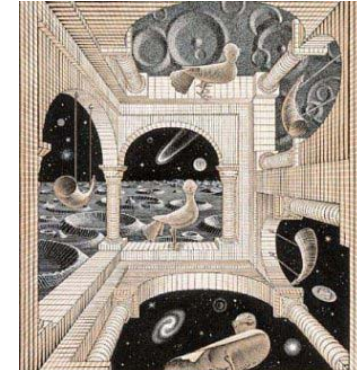
Science fiction authors and other futurists have long foretold of an age when we have become, for better or for worse, immersed in technology<sup>1</sup> to the point where information is available everywhere and at anytime. [3] [4] Reality has consistently caught-up with even the most outlandish predictions thanks to the continual and innovative development of technology, particularly with regards to cost, size, spatial integration, localization, interactivity, and communicative capability.[5] These advancements have led us to a world where our interaction with machines and information has become part of our environment and often indistinguishable from our natural lives. [6]

More recently, we have started to see these pervasive systems enhanced with media-rich capabilities such as the recording, transmission, and localized playback of high-quality audio and video. Whether we like it or not... we have arrived at *The Age of Ubiquitous Media*.

### 1.1.1 Ubiquitous Media Consumption

Traditionally, the concept of ubiquitous media is concerned with consumption. It is associated with the networked availability of content that provides us with endless access to information, education, and entertainment. This access is further supported by the proliferation of media-enabled mobile devices and theater-like outposts everywhere we go.

This increasingly wide selection of content and interactivity is assumed to provide a more personalized media experience. The availability of constant review information and other user-contributed analysis combined with increased competition amongst media producers and distribution channels further indicates an increase in overall quality.



<sup>1</sup> Jorge Luis Borges' short-story *Aleph*, [3] published in 1949, describes a point in space from where all other points in space and time become visible forming a theoretical basis for writers such as Verner Vinge [4] to herald the oncoming Ubiquitous Age.

More recently, audience measurement techniques have shown an increasing fragmentation<sup>2</sup> into niche markets. [7] Cited positive effects of demassification include:

<sup>2</sup>“At the most basic level fragmentation refers to the process wherein traditional “mass audiences” are subdivided into smaller, more homogeneous segments as a result of an increasingly segmented and complex media environment... Audience fragmentation refers to the extent to which media audiences are more widely distributed across a variety of content options.”

- Philip M. Napoli [8]

- 1 Playing field is leveled and democratized, lowering the barriers for entry allowing new players
- 2 Audience becomes grouped into small niche markets that can be reached without costly broad marketing
- 3 Choice of viewing method and level of interactivity provides a more individualized and engaging experience
- 4 Formation of niche communities promotes more detailed meta-content, which can vitalize a subculture
- 5 Increased competition leads to increased quality
- 6 Content becomes more contextual

Mass media research [10] has also identified a list of negative effects of demassification:

- 1 Smaller cultural groupings increase tunnelvision, self-replication, and can become anti-social, promoting a stagnant or limited world perspective. In other words, access to discordant ideas becomes restricted [9]
- 2 Creates a fragmented voice or millions of voices not relevant to each other or to society, ultimately “creating an inane and trivialized culture” [11]
- 3 Viewpoints illustrated in user generated content generally come from a small demographic who have the time and the means to produce unpaid content and analysis[11]
- 4 Media illiteracy results from untrustworthy content overwhelming and, in some cases, replacing professionally generated content

- 5 In order to reach a profitable marketshare, media producers and broadcasters must migrate horizontally, which requires larger marketing budgets, taking away from their development budgets
- 6 It becomes challenging to predict the market niche and the success of a product, causing a reliance on pre-sold franchises and non-novel content
- 7 As content becomes contextual and temporal, it also becomes perishable and difficult to assign ownership, resulting in failing profit models, potentially destructive to traditional, high-quality media

Despite these potential problems, user-generated content creation is rapidly increasing.

FIGURE 1-1 Current and predicted growth of UGC creation

**US User-Generated Content Creators, 2007-2012  
(millions and % of Internet users)**



As the quantity of unchecked content increases and the media-scape becomes more and more fragmented, the impact of these issues will become significant. We have already seen the effects on traditional media in the form of lost revenue, lay-offs, and shutdowns across print, film, and music. [13] In response to this threat, these industries have concentrated their efforts towards



Time Magazine voted “You” as the 2006 Person-of-the-Year romanticizing user generated content [12]

<sup>3</sup>Web 2.0 is the term, originally coined by Tim O'Reilly, that now represents web applications that enable interactive information sharing.[14]



<sup>4</sup>Marshall McLuhan, considered the father of media theory, analyzed the effects of various media on their audience including the anesthetization caused by overuse of a mechanism and the need to form a "global village" to prevent over fragmentation and a loss of the message. [18][19]

both traditional marketing and attempts to utilize the so-called Web 2.0<sup>3</sup> facilities to promote their products.

With its commercial intention, marketing material, regardless of its medium, is rarely culturally significant. [15] It quickly adds to the deluge of available information vying for our attention.

We are thus faced with a signal-to-noise problem from the endless user-generated content combining forces with the results of its competition with traditional long-form media obscuring the interesting, engaging, and personalized content that the ubiquitous media age promises. If information with limited cultural scope, questionable intentions, and a general lack of quality continues to fill up the bandwidth with noise, it can become too difficult for the end-user to find something of value. Currently, web search is enabled with textual, unvalidated information, mainly contributed by users or advertisers, and often provides unwanted and outdated results. This signal-to-noise problem can lead to a loss of participation that can ultimately cause a decrease in the production of all types of media. [16]

The research presented in this dissertation addresses these problems by developing tools, techniques, and theories for the creation of meaningful media-rich content that stays true to the tenets of the age of ubiquitous media, including free form expression and personalized experience. Inherent to this goal is identifying possible attributes that can signify the difference between worthwhile content that can have an impact on society and the fragmented information that can lead to an amateurish world ripe with media illiteracy<sup>4</sup>. These attributes can inform a system of tagging of media, allowing the consumer to best locate the signal in the noise.

### 1.1.2 Ubiquitous Media Production

This research is primarily concerned with video content since this medium was historically dominated by paid, long-form media providers due to its relatively high production and broadcast costs. Technological advances such as cheap cameras have shifted the power structure, allowing mass social engagement in participatory video creation. Media capture technology is exploding, and image/video/audio acquisition capability is pervading many of our common digital devices (e.g., laptops, mobile phones, appliances, toys, etc.) which are increasingly carried on our person, providing opportunities for even more personal media generation.

With little or no quality-based moderation preventing the broadcast of user-generated content, the barriers to entry for contribution of video are non-existent. Idealistically, this makes everyone a media participant with a voice and a means to document their individual life. With the added incentive of a potential reward of money and/or fame, the act of personally generating content has become increasingly popular and a major force in mass media. The massive publicity of user contributions that have received these rewards has created a problem of overjustification<sup>5</sup>. Recent studies have quantifiably shown the negative effects of overjustification on the quality of normally unpaid, user contributed creative content. [20]

This new form of ‘expression’ differs significantly from traditional media in that it has little concept of the audience. In the world of paid content, if the product does not reach enough of the target population or is of too low quality, it will financially fail. This relationship creates a gatekeeper of quality and stricter barriers to entry that are missing from most of the user generated video that is readily available for view today. Without these requirements of quality, there is no pressure for end-users of new media technology to improve their content’s significance and reach. In addition, the low barriers of entry



The Sony-Ericsson Satio phone can capture 12 megapixel images and HD quality video.

<sup>5</sup>Overjustification, coined by the psychologist Edward L. Deci, describes the problem that the presence of external rewards takes precedent over the desire to perform an activity out of the enjoyment derived from the activity itself. While extrinsic and intrinsic motivations can be present in any task, it is the salience of the external rewards that allow for overjustification. Deci, in his famous experiment, went on to show that performance is compromised by overjustification. [21]





Any camera, whether it is a traditional surveillance camera or the one on everyone's mobile device, that can capture information without permission is potentially invasive.

open the door for scammers and other untrustworthy contributors negatively affecting the aforementioned signal to noise problem.

The in-pocket and personally owned form factor of mass-available camera technology promotes image capture based on reaction, not on learned deliberate action. This limits the reach and significance of the captured image, for the most part, to the active participants involved in the original event. In other words, much of the content created in this way is intended as prosthetic memory with nostalgic/informative value only to those immediately involved. To the rest of the world, this appears as noise, often with no way of filtering it out. This leads to an increasingly narcissistic society, [22] furthering the divide between the audience and the creator.

The problems described above are accentuated by the issues surrounding ownership of images captured in this way. [23] The media information is physically possessed by the user of the camera, giving them complete control over its presentation and broadcast, regardless of the content contained in the image. The image can contain a likeness or other information that is proprietarily owned by the subject, not the creator, of the image. The notion of ownership is further complicated by the fact that the possessor of the information can modify it and select how it is broadcast, breeding resistance to these technologies due to concerns of privacy and security. Despite a difference in intention, constant recording and transmission ultimately become surveillance worse than the *Big Brother* from Orwell's 1984 [24] as it becomes intertwined in the fabric of society; an inescapable *Little Snitch*. [25] In addition, without defined ownership it becomes difficult to assign value and/or credit which can lead to future unavailability or loss of the most significant content.

It is the mission of this research program to investigate methods and develop example systems that support democratized creation of video content that avoids the potentially negative repercussions of the issues with the current state of the art as described above. Beyond addressing these concerns, it must

promote the creation of audience-significant media and/or provide a mechanism for the personally tailored filtering of information. In general, as David Brin has written in the *Transparent Society*, [26] systems must compensate with new functions and features for the losses of privacy and ownership. These research requirements are detailed in section 1.1.4.

### 1.1.3 Additional Attributes of the Ubiquitous Age

Besides the explosion of devices for capturing, storing, and broadcasting rich-media content, most visions of ubiquitous computing anticipate a world permeated by a dense sampling of sensors and endless sources of data. This is starting to arrive in the form of environmental sensors (motion, power, light) integrated into our surroundings and wearable (gestural, biometric, social) sensors integrated into our mobile devices. These additional streams of information provide enhanced access to human and spatial subjects which can be leveraged by the ubiquitous media creation facilities to enrich the content. The captured audio and video data can be labeled and constrained by this additional information so that higher-level meaning can be extracted from it. This tagging can happen at the time of capture supporting new camera features and sensor-derived tags can be attached to the video throughout its broadcast existence enabling better cataloging and searching. This is already happening with user-entered meta-data being attached to images/video and the formation of life blogs.

The prevalence of screens can also be an asset, as they can provide in-situ information and user-interface facilities to augment the creation process and the possible applications that can be developed.

In general, the heightened awareness from these pervasive systems leads to richer forms of communication, education, entertainment, and social/personal enlightenment.

### 1.1.4 Research Requirements

The following list contains the requirements of research in the field of ubiquitous media to address the current problems and still allow for the personalized experience of user-generated content.

- 1 Development of systems and methods must occur at the scale of society. The facilities of content creation must be commodified to truly democratize the process and provide the capabilities across all demographics. This requires a lessening of necessary resources such as time and money, achievable by integrating the tools into the environment or by exploiting existing sensing/computing resources already there and designing an intuitive or completely invisible user interface.
- 2 An applicable theory must be developed to differentiate between fragmented, insignificant content and content which is cohesive and meaningful for the intended human audience.
- 3 While an external reward is not impossible, the system must provide intrinsic motivation for its use.
- 4 The system needs to collect as much real-time information as it can from the environment and from the human subjects. This information can be analyzed using the theories developed in requirement 2 above to determine the nature of events.
- 5 The developed system needs to be distributed and multi-tiered. By placing devices where most appropriate, i.e. on-body for behavioral sensing, the highest level of detail can be achieved. Events that contribute to the cohesive overall meaning of a created piece of content are not tied to a location. To best capture an event, the effects it has on the surrounding locations and people also need to be captured. Therefore, the system must be pervasive and have a large area of coverage in social and physical space.

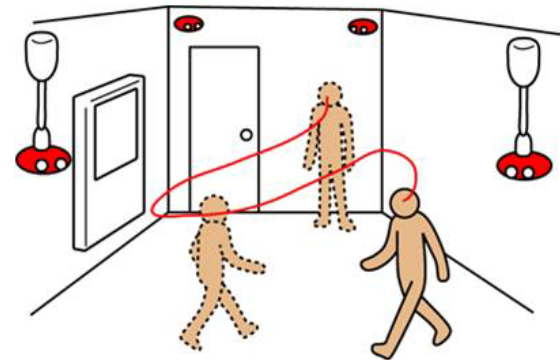
- 6 The system must not interfere with human behavior or be anti-social. In other words, it must be designed with the understanding that real human activity and social behavior are the best starting points for content that is meaningful to other humans. The ability to capture reality is the strongest attribute of user-generated content in comparison to traditional studio media. As much as possible, a system of these sorts must fade into the fabric of the environment.
- 7 Ownership, whether it is shared or unshared, must continually be assigned to all captured data. In-situ facilities for the immediate review of any recorded media by the owners is a desirable feature.
- 8 In-situ browsing of data and interacting with the system can also be used as feedback to personalize the media capturing and creation systems.
- 9 All captured and resultant information should be attached to the created media for browsing and searching.

The remainder of this chapter introduces the methods used in this research program to meet these requirements.

## 1.2 DISTRIBUTED MEDIA SYSTEM

The main requirement for research in the field of ubiquitous media is a platform for experimentation. At a minimum, this platform needs to provide the capabilities for the creation of user-generated video, i.e. it needs cameras, microphones, and networked accessibility.

Since the overall goal of this system is one of discovery, it needs to be completely modifiable both at the hardware level and at the software level.



By placing the creative technology into the environment, we assuage the antisocial nature of traditional content creation. This bases the output on reality and human social behavior, providing it with the potential for more appeal and significance, battling the fragmentation and media noise.

This will allow it to support many applications and foster endless experimentation across many research fields.

In order to provide as much information as possible to use to understand and label the collected audio and video, the platform must be equipped with as many sensing channels as currently available. It must also support application-specific sensor addons, and a device-to-device wireless communication network to connect to any source of information in the environment, such as the wearable sensors discussed in section 1.3, commercially available wireless sensors, and mobile phones.

The system's relationship with space and its existence as a facility for anyone to use are critical design factors. Most capture technology, particularly cameras and sensors, is a graft into the world and does not weave into the fabric of society. When dealing with the concept of creating content that is meaningful to the human audiences of society, it is essential that the tools integrate with but never interfere with the behavior of its subjects. This interference completely compromises and homogenizes the data, rendering the creation of meaningful output difficult.



Image from a SPINNER camera, purposefully blocked by a plant. One of the many forms of resistance to this type of technology.

It is also important to note that integration is not the same as invisibility. Hiding the cameras could create even more interference due to sinister implications of surveillance. The design of the individual devices and the overall deployment topology must walk a tight line to avoid physically overwhelming the environment and at the same time prevent an association with a surveillance system, resulting in resistance to its creation and use.

As with any distributed system, coverage is important, so pertinent phenomena are not missed, resulting in a loss of significant information. Actuation in distributed sensor systems allows the coverage to adapt to changing phenomena.[27] While mobile and personally owned cameras should be able to be integrated into the system to provide additional subjective views, providing a detailed third-person perspective into all the happenings in the

environment takes the system far beyond the current prevalent content creation mechanisms.

### 1.2.1 Obstacles to the Creation of this type of Platform

There exist major obstacles that have slowed the development of systems for the creation of video content at the scale of society. Currently, user-generated content is enabled by individual, non-systemic technological advancements in personal camera equipment and mobile devices.

One obstacle faced when attempting to develop an open-ended platform is the general complexity of video systems. Cameras and the computing power to capture, process, store, and transmit audio and video data are difficult to develop reliably. The cheap, integrated video solutions found in mobile devices, webcams, and surveillance cameras, do not provide the customizability and featureset required to support this research. It is essential to integrate the video capture with the additional sensors and enable user-created applications to exist throughout the system. The necessity of a completely custom system has precluded most, if not all, research in this field.

The major obstacle in the deployment of societal-scale video and sensor systems are the concerns of privacy and security. Since there is little or no real data concerning privacy and what people are willing to allow in exchange for new capabilities and experience, this issue can be seen as a research challenge and not an obstacle. An understanding of privacy concerns is becoming more and more important, since these types of systems will become commonplace. It is difficult to gather actual information about privacy in ubiquitously-sensed environments, since the systems that can investigate these questions cannot be deployed due to the very issue that they seek to explore.



<sup>6</sup> SPINNER stands for Sensate Pervasive Imaging Network for Narrative Extraction from Reality. The SPINNER logo, shown here, contains a branched network symbolizing connectedness and uses the snowflake to indicate the potential for unique individualized content that the ubiquitous age promises, but is backed by the radiation symbol indicating the dangers associated with a lack of research and understanding of this new media form.



## 1.2.2 SPINNER Media Platform

Evidently due to the obstacles listed above, existing distributed sensor and media systems such as those discussed in [Chapter 2](#) are based around a specific application and lack the sophistication and adaptability to fulfill the requirements of this research program.

To address this need, we present the first-of-its-kind SPINNER<sup>6</sup> distributed media platform. This system meets and exceeds the design requirements of a media-rich, societal-scale research facility and overcomes the obstacles that have historically interfered with this type of deployment. The SPINNER network of devices is permanently installed as part of the architectural space in the MIT Media Laboratory building, turning it into a responsive environment and enhancing life within and visits to its environs.

The devices in this system are input-output devices. In addition to their myriad of sensors, they also provide a small touch-screen display<sup>7</sup> and audio speaker. This way, information not only streams away from the user's environment – the system can also manifest virtual and remote phenomena into the user's space.

The SPINNER media network has been allowed to be deployed by addressing privacy issues in the following ways:

- 1 Provide an in-situ hard cutoff switch, disabling capture at any time
- 2 Approval for experiments running on the system has been granted by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES)
- 3 Utilize wearable and mobile technology to identify subject's privacy wishes to the system
- 4 A completely transparent hardware and software design, making available all source code and development tools



<sup>7</sup> SPINNER networked device, blending but not hiding in to the wall

5 Provide a suite of value-added applications to ease the system's uptake

6 Provide the system as a lab-wide research platform and installation

Fulfilling these points has allowed the system to be deployed. Further study of privacy conducted by Nan-Wei Gong has used the SPINNER media system to yield the first ever data set regarding the social acceptance of these types of systems. [28]

The SPINNER media system is discussed in greater detail in Chapter 3.

## 1.3 WEARABLE SENSORS

The tendrils of any sensate system must be placed where they are most effective. The networked media devices of the SPINNER system have been placed in the environment because they require architectural facilities such as high-speed networking, high-power for the motors, situated interface technologies, objective viewpoint, and complete area coverage. Locating the backbone of the system in this way allows the system to become a commodified facility focused on social reality.

However, the networked media devices do not provide enough access to the human subjects to enable the personalized data, social behavioral information, and individual interactivity to be used to create rich human-centric content and significantly categorize the captured media.

In order to address the above requirements, this requires on-body sensing. Although mobile devices are becoming increasingly sensor-enabled,





<sup>8</sup>The SPINNER  $\mu$ Badge provides outward looking social sensing information and audio recording



<sup>9</sup>Early prototype of the SPINNER wrist sensor unit provides gesture data and galvanic skin response

as of yet they do not provide the sensing and processing capabilities required for this research. They are also not positioned on the body in a useful manner. Therefore we have developed high-end, compact wearable devices that are tightly integrated with their subject, allowing access to data pertaining to identity, affective state, social behavior, and human gestural motion.

These wearable devices transmit detailed sensor data and identification information to the network camera devices tagging the recorded media. By providing the location and identity of the wearer captured by the camera, ownership of the image can be ascertained. The wearable devices are time synchronized to the cameras, and to any devices in the system, so that the data can be analyzed properly across multiple devices and for the synchronization of audio to video to sensor data.

The devices are in the form of a nametag badge<sup>8</sup> mainly focused on audio analysis and social interaction, and a wrist-worn device<sup>9</sup> mainly focused on gesture, physical state, and affective state. Both are equipped with enough processing power to classify the data in real-time should that be necessary. They also have OLED display screens that can be used for interactive applications or for the in-situ review of recorded video. The badge is also equipped with the ability to record high-quality audio directly from its wearer to be added to the video collected from the network, or to support distributed audio applications.

The SPINNER wearable sensor devices are discussed in more detail in chapter 4.

## 1.4 IDENTIFYING MEANINGFUL CONTENT

The main goal of this research is to suggest and test a possible theory that can elucidate the differences between meaningful cohesive content that a human audience would desire to view and fragmented chunks of media that provide little knowledge, significance, or entertainment to all but a miniscule audience.

We can then use this theory to create new tools and applications that are meaning-aware and can thus aid in the creation of significant human-centric content from the initial input. These techniques can also be used to label the audio and video data allowing browsing, searching, and filtering to be based on a meaningful metric.

Upon view of media, humans can intuitively decide what has personal meaning to them and what does not. The problem exists when there becomes too much information to ever view. This necessitates an automated system that can easily go through the massive amounts of data, and can behave in a human-like fashion to make decisions about the contents of the streams.

While this discussion begins to hint at an artificial intelligence problem requiring modeling of the human mind to form a general solution, the actual problem can be domain-constrained to that of the creation and communication of time-based content drawn from real events.

The challenge then becomes to identify and ultimately quantify what a human sees in a retelling of events that identifies one as a mere string of unconnected actions and another as a cohesively meaningful communication. This very question is the focus of the work of philosopher Paul Ricoeur.

Ricouer starts his discourse by identifying the concept of an event in our lives. He considers all events to exist only at the intersection of a sphere of



<sup>10</sup> Ricoeur's *Time and Narrative* [29] describes narratives as the most faithful articulations of human time. Narratives present the moments when agents, who are aware of their power to act, actually do so, and patients, those who are subject to being affected by actions, actually are affected. They also tell of worldly outcomes, intended or otherwise, of these interventions into processes that both predate them and outlast them. The historical time that narrative presents, i.e., human time, is an interpersonal, public time. It is the time in which one can locate sequences of generations and the traces their lives have left behind.

experience and a horizon of expectations. The sphere of experience is the collection of natural and cultural events remembered by the observer of the event, and the horizon of expectations is the unfolding of possibilities for the next event. Without these elements the event cannot be placed in cosmic time and lacks the actual substance and human meaning that creates the event for its subject in the first place. To convey an event, Ricoeur<sup>10</sup> claims, requires a discourse for not only communicating the strings of actions, but also their human contexts. He calls this discourse the narrative.

He goes on to define a narrative as having the following attributes:

- 1 Narratives organize disjointed elements into a cohesive unit, a plot that has a discrete temporal span
- 2 All events in the narrative appear as necessities to the plot, or at least appear to be related to one another
- 3 Narratives are not just a collection of actions, but they also contain characters with identities evident from within the narrative. The construction of the narrative, constructs the characters

In summary, Paul Ricoeur indicates that in order for a presentation of events to have any meaning, it must include a complete narrative discourse framing the sequence in human time with the above listed qualities. [29] This is in line with how humans naturally communicate.

Humans create stories as a method to understand fragmented events and to extract and communicate meaning. The narratives that we develop provide insight into our lives and the world around us, and are essential mechanisms to our education, entertainment, and social well-being. [30] Or as Toni Morrison said "[Narrative] is, I believe, one of the principle ways in which we absorb knowledge." [31]

The study of the conveyance and the content of narratives is called narratology. The Russian Formalist school began to study Narratology in the

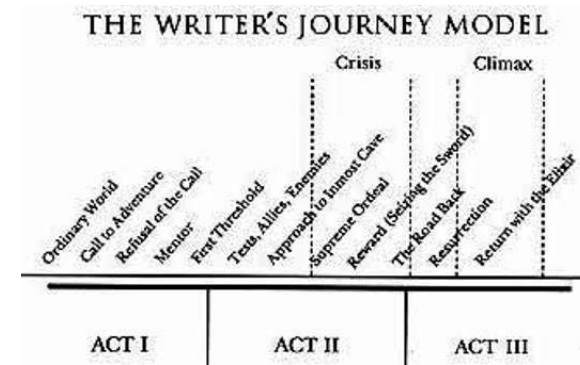
1920s. They divided the composition of a story into the *fabula* and the *syuzhet*. The *fabula* contains the story's themes, characters, and main points, and the *syuzhet* is the artistic and syntactic structure by which the *fabula* is conveyed. This school, as characterized by Vladimir Propp's work with Russian folktales [32] first suggested that all traditional story forms mix and match *fabula* fragments from a finite set of available scripts such as "hero saves damsel" all of which can be further reduced to a set of thirty-one actions that construct the narrative.

This was an early attempt at trying to understand the underlying requirements that distinguish a cohesive narrative from just a series of events. In Propp's case, he suggested that narratives require very specific archetypal elements to cohesively convey their morals.

Traditional long-form content is almost always a form of storytelling. Screenwriter Blake Snyder has developed a system for creating cohesive, narrative screenplays using a formula similar to Propp. His Beat Sheet 2 method identifies specific elements that must occur in a specific order to properly and completely convey the story. [33]

While it may be difficult to use sensors to identify all of the elements of these complex narrative models, the techniques used by narratologists can inform the mapping of captured events into cohesive narratives. These techniques become particularly useful when combined with ideas of standard narrative structure such as the 3-Act story<sup>11</sup> and the loss-search-find-change model.[34]

The use of these narrative concepts as a means to understand collected data and develop tagging systems for captured media is shown in Chapter 5.



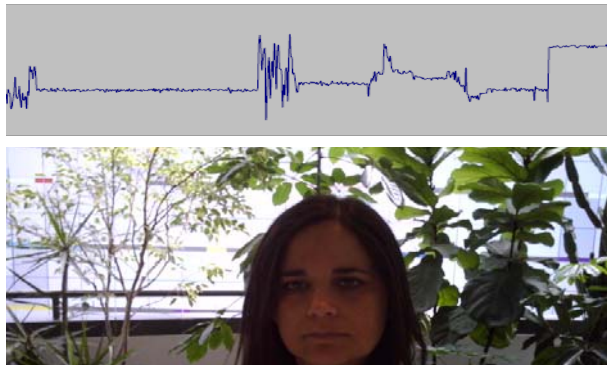
<sup>11</sup> Vogler's Writer's Journey model relates Propp's morphology to standard 3-act structure [35]

## 1.5 INTRODUCTION TO APPLICATIONS

In order to evaluate the concepts and systems innovated for this research program, a series of test applications have been developed. These applications are introduced here and presented in full detail with evaluated results in Chapters 6 and 7.

### 1.5.1 Labeling and Browsing

By providing additional streams of data and methods to understand this data, we can classify and label the captured audio and video content. To support this application, a wireless time synchronization system was created and resides in all devices. This system allows all time-based data collected across the network, to be in constant sync. A backend relationship database organizes and stores all the collected audio, video, and sensor data.



Sensor data is continually being capture in sync with the video data supporting many applications

Examples of types of raw sensor data that are collected along with the video are arm motion, body motion, galvanic skin response, room activity, location, and a person's rotational orientation with respect to the earth's magnetic field. This data is passed through algorithms that map it to higher-level information such as conversational dynamics, audio affect, and attention. The raw data and the mapped higher-level information are then passed through a classification system that can look for specific events, such as narrative events (a.k.a. story beats) or specific gestures. This information is then fed back into the database allowing the search of the media using any available information.

### 1.5.2 SPINNER Application

The main SPINNER application is the original motivation for research in this field. This application allows a user to input information about what type

of video content they would like to create. This information is ultimately equated with a higher-level story model, essentially an ordered list of the types of actions that are important and the characters contained in the desired output.

Using the database of audio, video, raw sensor data, mapped sensor features, and classified story beats as described in the previous application, the system automatically edits a cohesive video according to the user's parameters. In other words, a user-specified narrative discourse is created by selecting and combining the recorded, fragmented events of human behavior at the societal scale.

This SPINNER application is at the forefront of content creation in the age of ubiquitous media. It serves as a demo-able application, and it also serves as an experimental framework for research in the very nature of content creation, ubiquitous media, and applications of narratology.

The SPINNER application enables an entirely new form of entertainment, where participants watch events from their life assembled into a cohesive video that conveys their desired meaning in the form of a comprehensible narrative discourse. The created narrative video can serve as pure entertainment or provide new insights into one's self and society. The exercise of creating a narrative discourse from one's natural behavior can provide for its users new perspectives into the everyday by metaphorically linking real activities and previously non-apparent relationships with a narrative. The resultant video can be seen as a new form of diary/blog and multiple users can input the same parameters to see how their video differs, leading to the creation of a community.

### 1.5.3 Searching the Future

Since the data becomes available immediately, the system can be told to report on a particular condition the moment it happens. For example, the system can search for a handshake between Person A and Person B, when



SPINNER users wearing badges and wrist sensor units are recorded shaking hands by the SPINNER camera system. This clip is easily identified by the subjects' identities, location, and gestures provided by the wearable sensors



Person A recently talked to Person C with a ratio of talking of 10:1 in favor of Person C. Whenever this sequence of events takes place, the system can record it and send it to the query creator or create any type of notification necessary.

### 1.5.4 Sensor-Augmented Photography

As in all of the applications, the sensor data can be used to enable and disable the camera so that it captures only imagery that is desired and to allow the creation of pretriggers so that the condition of interest is not missed. In addition, the sensor data can be used to change the way in which the image is captured. For example, the light level combined with the motion detected by the wearable sensors can be used to determine aperture and exposure time to minimize subject blur and grain. Other conditions might call for a softer focus or changes in the depth of field, zoom, or camera angle as all the cameras are equipped with motors. Camera angles are often used to highlight aspects of the story in the current scene. An example of this is to have the camera lead the actor when the character is moving quickly increasing the appearance of motion, and when a character is failing in some way the camera will look down on them, diminishing their appearance. The SPINNER cameras can move in reaction to the social and behavioral data collected from the wearable sensors.



A SPINNER user is reviewing and approving a captured video clip.

### 1.5.5 Visualization

The SPINNER media network integrates real-time sound, images, and diverse data sampled throughout an environment that can be transformed into a visual, aural and haptic experience. These types of visualizations allow the observer to sense the overall activity of the area surrounding the system. The visualization applications can be interactive and the user can actively decide what information from which locations they would like to have visualized. These installations reflect the essence of the physical environs as a holistic entity.

### 1.5.6 Situated Interactive Applications

Since the SPINNER devices are equipped with a touchscreen we have created several applications for use in-situ. These include:

- 1 In-situ browsing of a sensor data labeled map of the building to determine where to go
- 2 Location-to-location audio/video streaming
- 3 Display of personalized message content
- 4 Use as a distributed display to play location and identity based media content
- 5 Multipoint still image capture and broadcast system
- 6 Interaction with virtual environments

Applications such as these are essential as a value-added services to aid in the acceptance of the system.



<sup>12</sup> SPINNER node running the multipoint image capture and broadcast application



# PROLOGUE

.....

.....

Given that external reality is a fiction, the writer's role is almost superfluous. He does not need to invent the fiction because it is already there.

- JG Ballard [36]

The *Prologue* chapter contains the related work to frame this research in the larger field.

## 2.1 RELATED WORK

The overall research program presented in this document has no direct precedent due to its novel and complex application space supported by a combination of a situated distributed video network with wearable sensing and its use of narratology to philosophical ground its internal methodologies. However, there has been much work in the related fields from which sub-components of this project have grown.

### 2.1.1 Video Sensor Networks

Wireless sensor networks have become a large area of research, with many universities and institutes contributing. Strategic seed programs begun in the 1990s such as DARPA's SENSIT initiative [37], have grown into an international research movement.

Advances in this field have lead to smaller, cheaper sensor nodes[38]. The SenseCam<sup>13</sup>, developed at Microsoft research [39] and now being produced by Vicon [40], brings video sensing and image gathering to a small, low-power node. Devices such as this have started to support a number of distributed camera systems such as the ones developed by Wayne Wolf at Princeton as a test platform for distributed vision algorithms [41]. The Panoptes system [42] developed at OHSU shows how redundancy in video camera systems can keep the information detail even in the event of a network outage as well as demonstrated a reprogrammable video platform.

The work of Andreas Savvides' lab at Yale, called Enalab, often uses multiple camera systems to investigate phenomena and the real-time information captured [42], and often uses additional sensing capabilities for subject identification and labeling of video. [43]



<sup>13</sup>The MicroSoft Research SenseCam [39] constantly records image data acting as prosthetic memory

While not exactly a distributed video network, Microsoft research's MyLifeBits [49] works to address challenges implicit in organizing burgeoning streams of heterogeneous digitized personal media. Systems of this type would be quite complementary to a system like the SPINNER video network.

Deb Estrin at the CENS lab at UCLA has presented projects concerned with the power-usage, coverage, [44] and effectiveness of distributed camera networks. [45] The effectiveness of distributed media systems for interactive environments has been investigated by designers such as Scott Snibbe and Hayes Raffle. [46]

An invasive, pervasive system of sensors was designed for a privacy study conducted at EuroParc. This system helps to provide some prior experience on acceptability of these types of systems. [47]

### 2.1.2 Human-centric Sensing

Advances in ubiquitous and wearable sensing have allowed us to observe human subjects without interference. This supports investigations into organizational behavior, social networking, and information diffusion. The Human Dynamics Group at the MIT Media Lab has been a leader in this area through several projects that look at human social behavior, such as the Sociometer [50] and the Reality Mining project [51]. Recent collaborations with the Responsive Environments group have utilized the Uber-Badge<sup>14</sup> platform to look at group social signaling and interest [52]. This has led to a new platform called the Wireless Communicator [53].

Research into the mapping of sensor data to human behavior is exemplified by projects such as Singh's LifeNet [54], Weld's work in personalized user interfaces [55], and Wolf's work on understanding the purpose of travel from GPS data [56]. More specifically, the field of activity recognition from wearable sensors, such as the work of Stephen Intille at MIT that can classify activity from accelerometer data to provide context in



<sup>14</sup>The Uber-Badge system mapped sensor data to “interest” in social situations

healthcare applications and Paul Luckowicz's work with ETH and the University of Passau on activity recognition using a suite of wearable devices [58], has provided insight for the design of new wearable systems. There is also a rich body of literature relevant to designing a mapping of this sort including Ro et al's Pattern Classification text [59] and Wertheim's discussion of organizational behavior [60].

In Bove and Mallett's paper "Collaborative knowledge building by smart sensors" they explore "decentralized approaches for gathering knowledge from sensing devices" [61]. This body of research is perhaps the closest relative of the SPINNER system as it attempts to identify knowledge in the data collected from reality. It discusses several examples, including the apropos Two Eyes camera project [62]. In a similar vein, Bo Morgan's research is concerned with the use of commonsense algorithms for understanding sensor network data [63].

### 2.1.3 Meaningful Content vs. Discordant Events

The most important body of work to examine when discussing what makes certain content worth consuming is the entire world of well-written, well-crafted, traditional content. Of particular note are authors such as David Rakoff [64] whose style transforms an almost journalistic approach based on first-hand experience into an insightful, humorous, and entertaining story that is almost always an improvement from the experience itself.

Investigating the techniques by which this happens in the job of the narratologist as introduced in section 1.4 via references to Ricoeur [29], Propp [32], and Kearney [30]. Story structure<sup>15</sup> is an important topic for screenwriters [65] to help develop effective narratives. The field of story generation looks into developing parametric models of narrative to create systems for interactive entertainment. An example of this is Meehan's Soap Opera plot



<sup>15</sup>“Steal This Plot” by June and William Noble [65] is one of many texts discussing a finite set of plot formulas that can be used to create a narrative. This idea is summarized in the Blake Snyder mantra, “Give me the same, but different!” [33]

generator [66]. Prior to this, in 1973, Klein's essay discussed methods for automatically writing a novel [67]. Glorianna Davenport has continually performed research into our use of stories for interactivity and understanding [68] and her Media Fabrics group at the MIT Media Lab has been using narrative ideas as a means for communication [69] and for looking at the design of and our relationships with everyday objects [70].

Using story concepts to understand, catalog, and browse information has been discussed as far back as 1977 with Schank et al's book "Scripts, Plans, Goals, and Understanding" [71] followed in 1982 by Lehnert's article [72] describing the use of plot units and narrative to summarize text. Andrew Gordon's PhD thesis [73] discussed using additional knowledge such as narrative structures, common-sense, and sensor data to browse information. More recently, the StoryNet project [74] has built a database of scripts that are used to search through and predict subsequent elements of large collections of data.



# SYSTEM DESIGN

A decorative dotted line consisting of a horizontal row of dots followed by a vertical column of dots, forming an L-shape.

Since I see technology as being an extension of the human body, it's inevitable that it should come home to roost.

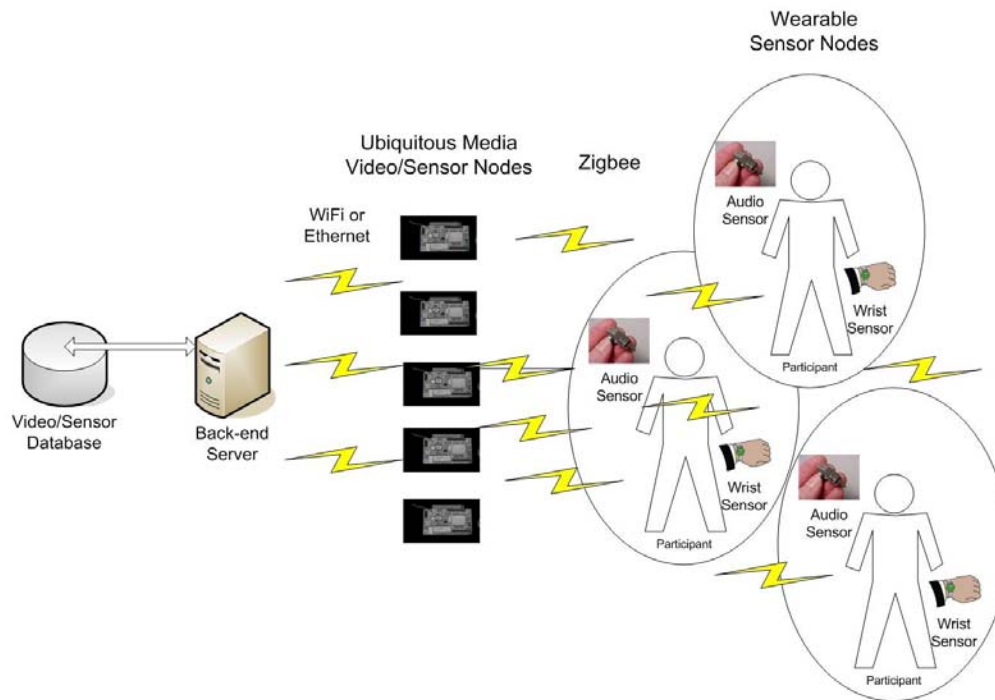
- David Cronenberg [76]

This chapter contains the technical details of the system that was designed and built as a platform to support research in ubiquitous content creation and understanding.

## 3.1 OVERALL SYSTEM ARCHITECTURE

This research sits at the intersection of several fields that encompass pervasive sensor networks, wearable computing, and distributed video systems. Currently, all research platforms and sensing products concentrate on a single space, either wearable, architectural, or social, and generally focus on a narrow range of applications. To enable research that leverages a broader range of capabilities across many fields, a novel multi-tiered platform has been designed, built, and deployed.

FIGURE 3-2 Multi-tiered system





## 3.2 DISTRIBUTED MEDIA SYSTEM

The heart of the multi-tier platform is the SPINNER distributed media and sensor system. It integrates into the environment and provides situated, always-on, high-powered, high-bandwidth capabilities to any application. The current deployment contains 50 individual nodes, sometimes known as Sensate Ubiquitous Media Portals, installed throughout The Media Laboratory complex at the Massachusetts Institute of Technology. Each of these *portals* is built from a collection of modular components.

FIGURE 3-3 Sensate Ubiquitous Media Portal

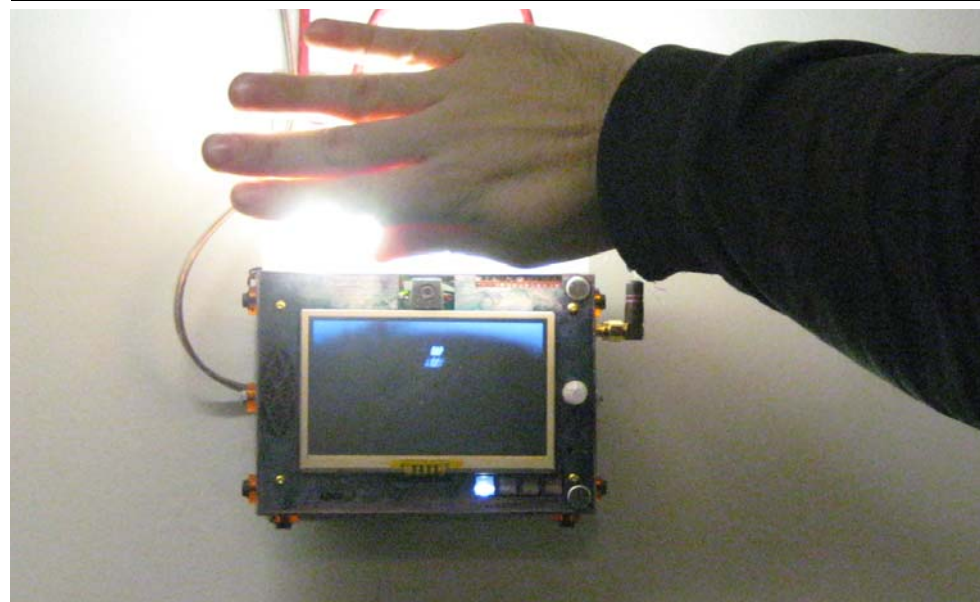
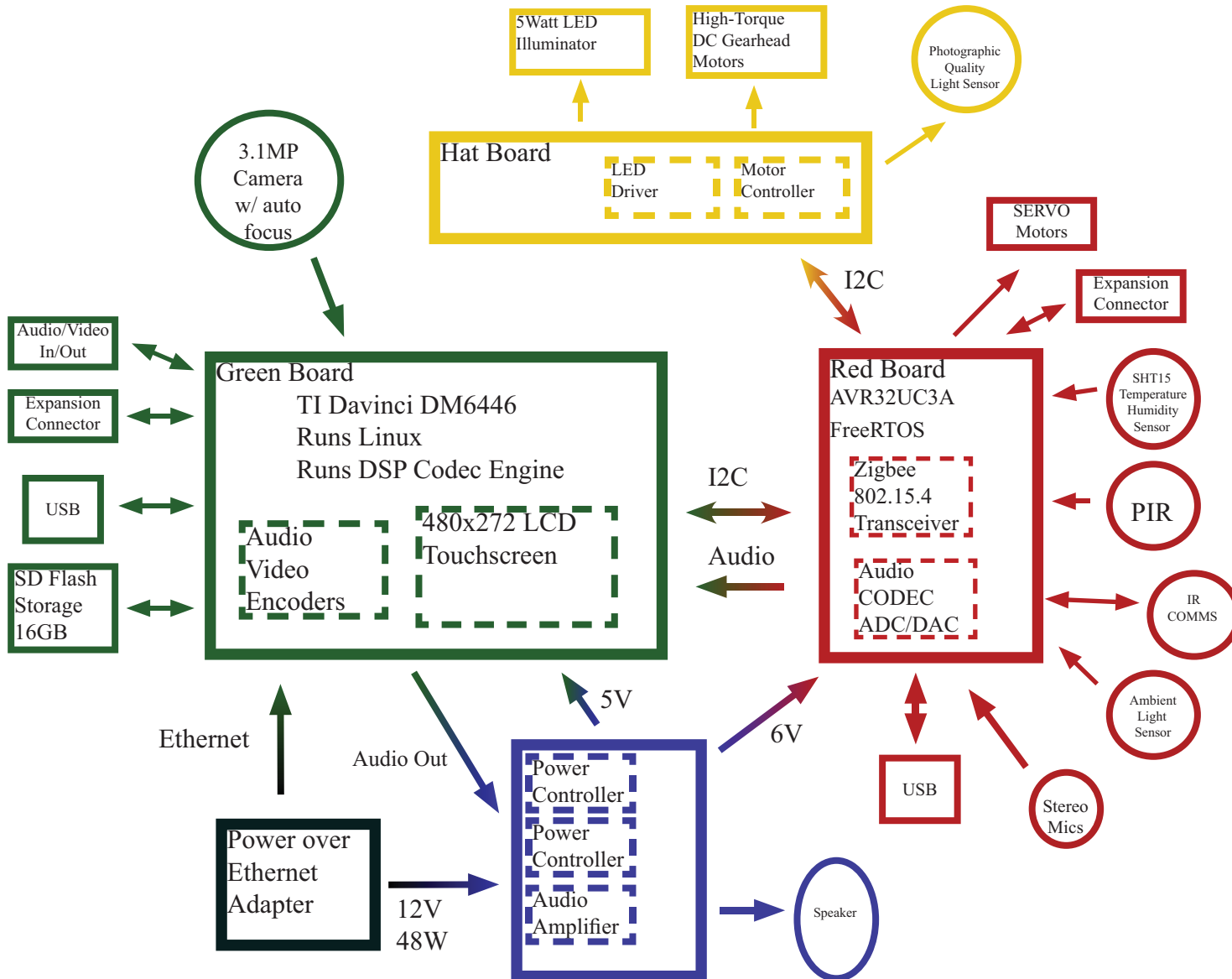


FIGURE 3-4 Ubiquitous Media Portal Block Diagram



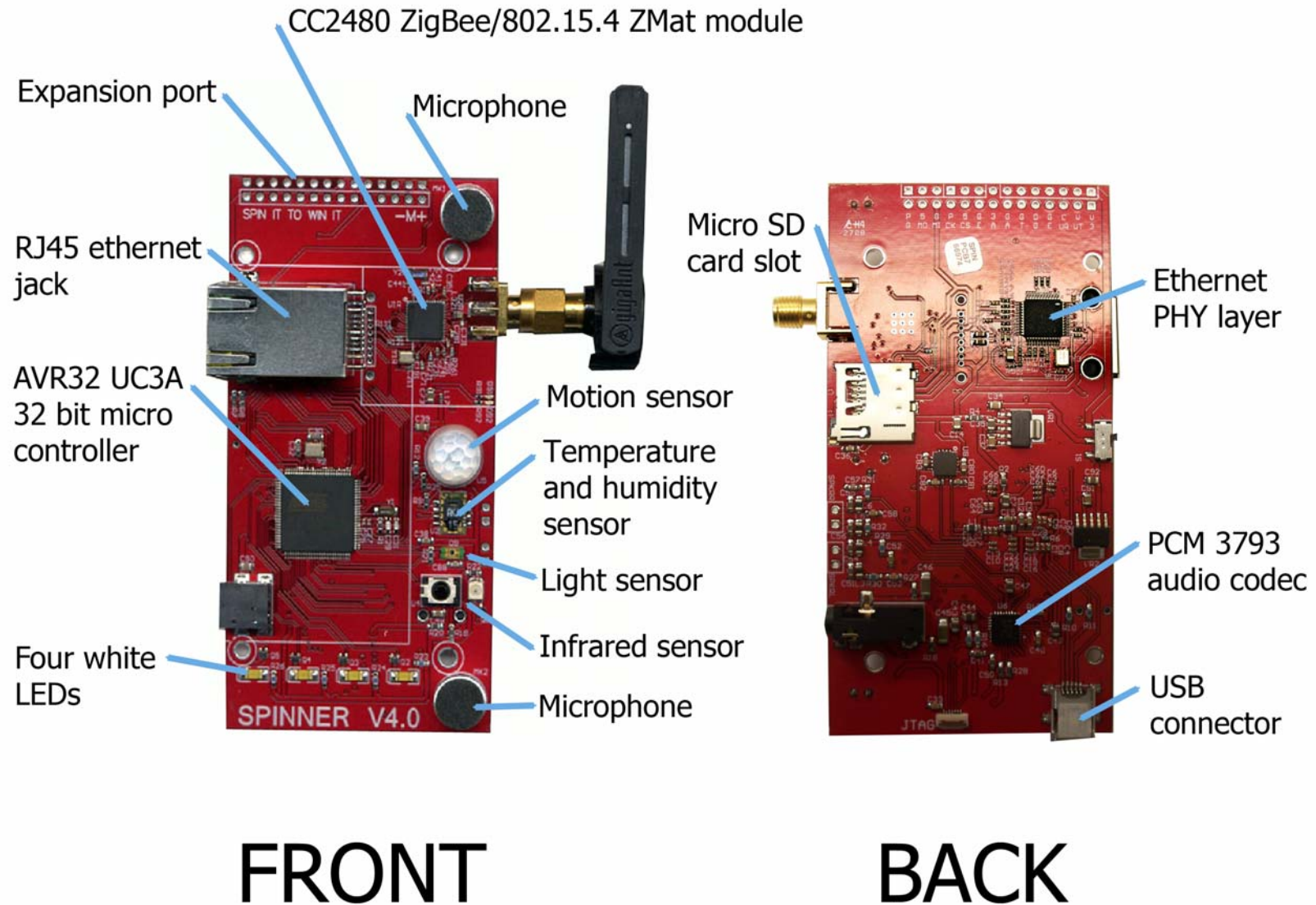
### 3.2.1 Modular Component #1 - The Red Board

The first component of the distributed sensor/media system provides the room-level sensors, audio capture system, and infrared communications. This component, nicknamed The Red Board for obvious reasons, also contains the ZMat wireless module comprised of a Chipcon Zigbee/802.15.4 radio chip, proprietary software extending the features of Zigbee, and the required analog components. For the Red Board, the ZMat module is set to act as a ZigBee router. The Red Boards form a wireless mesh network covering the entire complex allowing any mobile devices or other sensors to be discovered and communicate through the network as a backbone. The ZMat transceiver also acts as a beacon for a basic location system. For higher bandwidth communication and control, the Red Boards also have embedded ethernet, providing internet access to every device in the system. The Red Board can be combined with other modules to form a Ubiquitous Media Portal, work completely standalone with its embedded ethernet, or be combined with any other device in order to integrate that device into the network, giving it access to all the features of the system.

#### 3.2.1.1 Red Board Sensors

- 1 PIR Sensor - The Panasonic AMN11111 Passive Infrared Sensor provides an accurate assessment of whether there are moving human bodies in the room. The field of view of this sensor has been selected to be similar to that of the camera in the Ubiquitous Sensor Portal video system
- 2 Temperature and Humidity Sensor - The SHT15 is the industry standard factory calibrated temperature and humidity sensor
- 3 Ambient Light Sensor
- 4 3-axis Accelerometer - Used to detect room/wall vibration or to be used for knock detection as a mechanism for user input

FIGURE 3-5 Component diagram of situated sensor node, a.k.a. The Red Board



### 3.2.1.2 Red Board Audio System

The audio system on the Red Board utilizes a Texas Instruments PCM3793 audio CODEC chip. This chip provides phantom power to two electret microphones, up to 50dB of amplification, automatic gain control, filtering, and analog to digital conversion. It can sample the two microphones with a bit-width of 16-bits per sample at a sampling rate of up to 48,000 samples per second. The firmware in the Red Board defaults to 12,096 samples per second, which is adequate for speech applications, and all aspects of the audio system are user-adjustable through any external control interface that the Red Board supports. The audio system can also be set into a pass-through mode where the analog features of the CODEC are utilized, but instead of converting it to a digital signal, it sends the clean, amplified audio signal out as analog.

### 3.2.1.3 Red Board Processor

The main processor on the Red Board is the Atmel AVR32 UC3A. This processor is a 32-bit microcontroller with DSP instructions. It runs at 66MHz and has 512Kbytes of flash to store programs and data. The AVR32 has 64KBytes of dual-port SRAM and a suite of peripherals, making it ideal for this application.

### 3.2.1.4 Red Board ZMat Transceiver

The ZMat wireless transceiver is the heart of the entire system. It forms a mesh network with all of the always-on and listening Red Boards. This network allows all devices in the system to communicate peer-to-peer as well as to centralized services. Since the Red Boards are also equipped with ethernet, the wireless transceiver provides internet and high-bandwidth

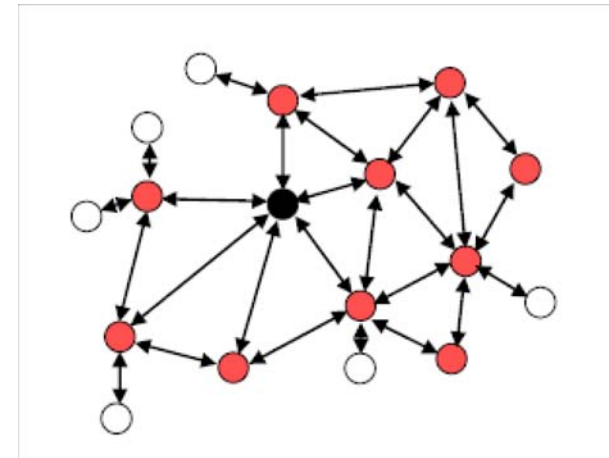
network services to mobile, battery-powered devices that only have a low-powered wireless transceiver.

The ZMat wireless transceiver module is based on the Texas Instruments/Chipcon CC2480 Zigbee Applications Processor. This processor is completely compliant with the Zigbee/802.15.4 specification, allowing it to communicate with any Zigbee device that can support a custom Zigbee Application Profile. The ZMat transceiver module on the Red Board is set up as an always-powered, always-listening basestation. Upon board power-up, the ZMat transceiver searches for an existing ZMat Zigbee network to join. If it finds one, it becomes part of the mesh network as a router/basestation and maintains a routing table. If it does not find an existing network to join, it starts a new network, allowing other devices to join to it and start to form a new mesh. Should this mesh eventually overlap with an existing mesh, the two networks can join into one. If the individual mesh networks never overlap wirelessly, they can be joined using the backend wired network.

The CC2480 chipset allows a user application to access the underlying radio functions while still running a ZigBee application. The ZMat software takes advantage of this and adds peer-to-peer data streaming at 150kbps, enough bandwidth for 20 channels of ADPCM encoded speech audio or high-speed download of stored data and firmware. In addition, the ZMat firmware collects RF signal strength and digital quality information from every communication received.

Mobile devices equipped with a ZMat transceiver setup as a low-powered ZigBee End-Device are identified, registered, and continually tracked by the Red Board mesh network. These end-devices can then transmit data to anywhere in the network and receive clock synchronization and location information from the basestations.

**FIGURE 3-6** The ZigBee mesh network topology containing a single coordinator, a multitude of routers, and mobile end-devices.





### 3.2.1.5 Red Board Infrared System

The Red Board is equipped with an IR communication channel (875 nm modulated at 38 kHz) to support line-of-sight communication. Devices can notice each other via this IR channel out to 3 meters and across large angles (e.g., 60°). All devices sporting this IR system broadcast a packet containing a unique ID code through their IR port to alert other nearby facing devices of their presence. Although the average interval between IR pulses is 1 second, it varies by up to 25% from shot to shot to avoid persistent collisions.

In the case of the situated Red Board, the IR channel is mainly used to identify a badge-user who is interacting with the system. The IR channel can also be used as a location beacon and as a more private than RF in-situ communication channel.

The Red Board also has an IRDA interface for communicating with a PDA or other computing peripheral. As an example, the IRDA interface was used specifically for communication with the MIT Media Lab Human Dynamics Group's social communicator badge [53].

### 3.2.1.6 Red Board Wired Interfaces

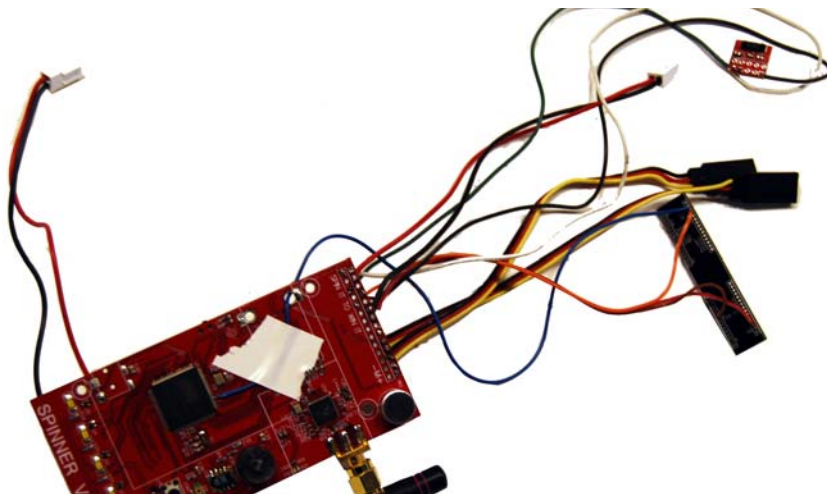
The Red Board supports several wired communication channels for integration with other devices and/or existing networks. All sensor data and network packets collected on the Red Board are made available simultaneously on the following wired ports:

- 1 Ethernet - The Red Board has an Ethernet PHY chip and uses the LWIP TCP/IP embedded network stack. Client applications can hook directly up to the Red Board via a TCP/IP sockets connection and receive streaming data. The Ethernet jack can easily be removed for applications that do not require it
- 2 USB - The Red Board can act as either a USB peripheral or USB host and stream data to a connected device

- 3 UART Serial - A standard wired serial port operating at 230400 bits per second is also available
- 4 Synchronous Serial - Master and Slave versions of both I2C and SPI are available for external devices to communicate with the Red Board. The I2C operates at 400KHz and the SPI operates up to 25MHz
- 5 Expansion port - The Red Board supports the direct addition of hardware peripherals by providing a multitude of analog and digital I/O channels for expansion
- 6 SD Card slot - In addition to transmitting all of the collected data, the Red Board can record it using a FAT32 filesystem to an onboard microSD flash memory card

Client devices and applications can attach through any of these ports and control the operation of the Red Board, minimizing the need for firmware modification.

**FIGURE 3-7** The Red Board with flying wires, ready to be integrated into a Ubiquitous Sensor Portal





### 3.2.1.7 Red Board Firmware

The Red Board's firmware is designed to be scalable and easy to add additional features by anyone, regardless of their knowledge of the intricacies of the hardware. However, the firmware is designed to be completely controllable during run-time and requires no code modifications unless hardware is added to the Red Board's expansion ports.

In order to achieve this design goal, the Red Board firmware uses FreeRTOS at its core. FreeRTOS is a very small footprint task scheduler allowing the firmware to be divided into individual tasks that can easily be added, removed, enabled, and disabled according to the application and the operating conditions. FreeRTOS provides a layer for the tasks to communicate with each other and maintains individual stacks and program counters for each task.

The Red Board firmware performs the following functions:

- 1 Wraps all the peripherals of the microcontroller with OS-level tasks, allowing easy, queued operating of all hardware functions
- 2 Provides an abstracted interface for all communication channels so the firmware can write one packet and have it sent on multiple channels according to the current settings
- 3 Implement standards such as TCP/IP, FAT32, and the ZigBee application layer
- 4 Scan all sensors. The firmware defaults to 100Hz sampling rate for all of the sensors. This sampling rate is adjustable at run-time with commands sent to the Red Board through any communication channel. Changes to these settings are persistently stored in non-volatile memory
- 5 Listen on all wireless communication channels

- 6 Send periodic pings on all wireless channels, enabling the location system and identification system
- 7 Operate the audio CODEC and receive digitized streaming audio recorded from the microphones on a high-speed I2S bus. It can then compress and store this audio in a large circular buffer. Client applications are notified when audio is captured so that the audio can be retrieved. All audio CODEC parameters are modifiable by the client applications
- 8 Implements optional filters on the incoming sensor data
- 9 Makes all sensor data, incoming communications, and state digests available to client applications on any communications channel
- 10 Drive two servo motors
- 11 Control the brightness of the four LEDs and provide status indication through LED animations

The Red Board has been designed and has been operating as a turn-key sensor network system that can easily be integrated into devices, larger systems, and the environment.

### 3.2.2 Modular Component #2 - The Green Board

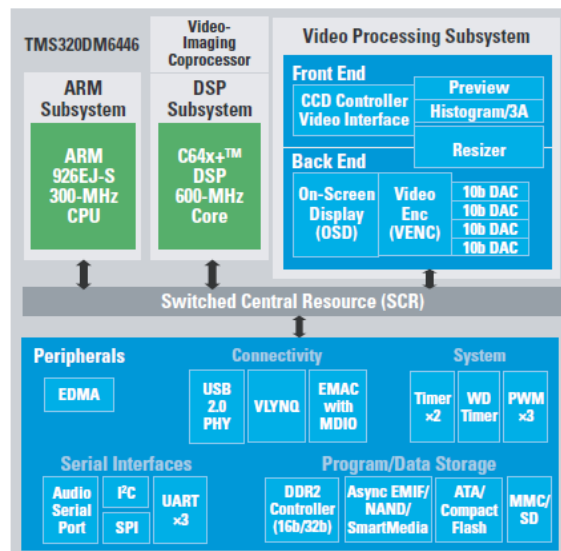
The next module that the system requires provides the media features, such as video capture, and the processing capabilities to handle high-bandwidth data. In addition to providing these advanced features, this module runs a standard operating system to take advantage of existing development tools and open-source software. This module, nicknamed The Green Board, adds all the features of a full PC to each node in the distributed system. Unlike a full PC, however, the Green Board is designed to minimize its footprint to aid in its deployment, operates with precision timing and real-time features, and

provides a suite of hardware features for multimedia capture, processing, and display.

### 3.2.2.1 The Green Board - Hardware Design

The Texas Instruments' Davinci series of microprocessors was selected as the main CPU of this device. More specifically, the Davinci DM6446 was an ideal fit, as it is a dual-core solution with a 600MHz ARM core to run the OS and user applications, a 1GHz 64-bit floating DSP core for advanced video processing, and many hardware features dedicated to video such as high-speed interfaces to cameras and LCDs, a hardware resizer peripheral, and a shared memory bus for the ARM, the DSP, and the video peripherals.

FIGURE 3-8 Davinci Block Diagram [77]



Texas Instruments graciously donated a complete development setup, including an evaluation board, complete software development kit, production-ready audio and video CODECs that run directly on the DSP core of the Davinci chip, and one year's worth of technical support. Texas Instruments also provided the contact information for Empower Technologies, a manufacturer of single-board computers using the Davinci series microprocessors.

Working with Empower Technologies, a complete embedded Linux video solution was developed, including a Davinci-based single-board computer, a 3.1MP camera module, a 480x272 24-bit RGB touchscreen LCD module, and an open-source Linux operating system and toolchain.

FIGURE 3-9 Front side of the Green Board

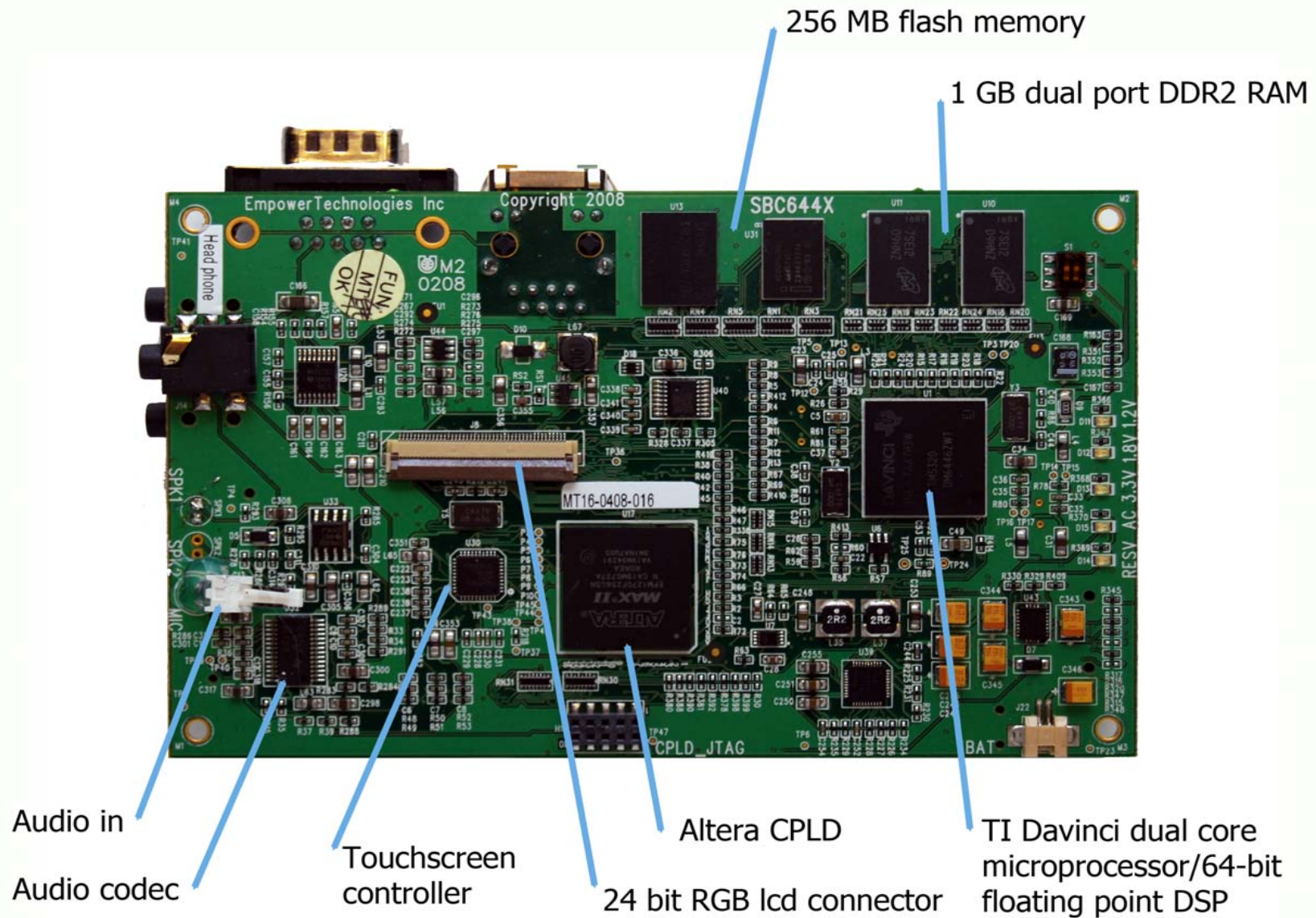
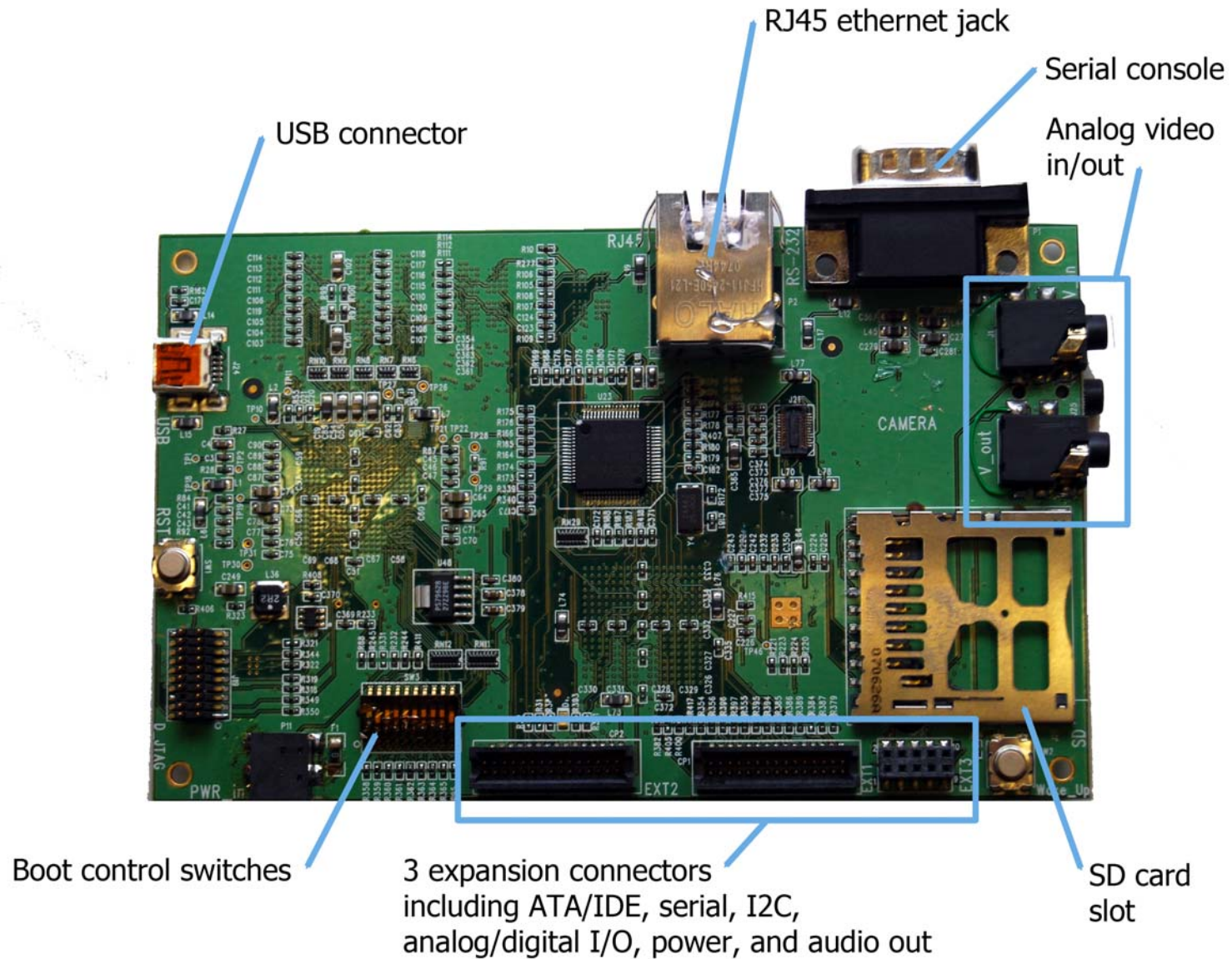




FIGURE 3-10 Rear Side of the Green Board



**FIGURE 3-11** The Camera Module on its custom cable



The camera unit has a motorized auto-focus and a 16-bit parallel high-speed digital interface. The Davinci processor has a dedicated peripheral that can capture from the camera module directly into RAM. This memory is then accessible by the DSP core for processing and by the hardware video peripherals, such as the resizer or the previewer, which can send the video to a display device without processing.

The camera required a custom cable assembly to allow it to be mounted on a panning motor or anywhere in the final package. This cable has particular design requirements to support the high-speed digital data. At each end of the cable is a custom PCB that adapts the cable to the correct connector and contains circuitry to buffer the signals.

### **3.2.2.2 The Green Board - Operating System**

The Davinci community maintains a GIT repository containing a Linux kernel source tree. This kernel is based on the basic ARM Linux kernel with modifications to support the Davinci specific hardware features. In addition, Empower Technologies provided modifications and additional kernel module drivers to support the particular peripherals included on the single board computer. All of the features of the hardware, such as sound, video display, touchscreen, networking, camera support, and wired expansion ports, are wrapped by standard Linux drivers and shared libraries to allow standardized access from user applications. All of the features of the Davinci chip itself are also abstracted and made available for an application to use with standard C

**FIGURE 3-12 Davinci Linux kernel source GIT tree**

<a href="#">summary</a>	<a href="#">shortlog</a>	<a href="#">log</a>	<a href="#">commit</a>	<a href="#">commitdiff</a>	<a href="#">tree</a>
<a href="#">snapshot</a>					
rebuild linux-davinci from branches					
master					
-rw-r--r--	.gitignore	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	.mailmap	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	COPYING	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	CREDITS	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
drwxr-xr-x	Documentation	<a href="#">tree</a>	<a href="#">history</a>		
-rw-r--r--	Kbuild	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	MAINTAINERS	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	Makefile	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	README	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
-rw-r--r--	REPORTING-BUGS	<a href="#">blob</a>	<a href="#">history</a>	<a href="#">raw</a>	
drwxr-xr-x	arch	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	block	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	crypto	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	drivers	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	firmware	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	fs	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	include	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	init	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	ipc	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	kernel	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	lib	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	mm	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	net	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	samples	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	scripts	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	security	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	sound	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	tools	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	usr	<a href="#">tree</a>	<a href="#">history</a>		
drwxr-xr-x	virt	<a href="#">tree</a>	<a href="#">history</a>		

code examples. Beyond what is required to support the custom hardware, the SPINNER system required the modification of the kernel to support high resolution timers, preemptive scheduling, and full-duplex audio.

A Crosstool-generated GNU toolchain is provided for compiling user applications, kernel code, and open source programs and tools. This toolchain can be used directly on a device running the Davinci Linux OS, or can be used on any Linux host to cross-compile programs for deployment on the devices. These standard tools can be used with any IDE, such as Eclipse, and with the GDB debugger.

Many open source Linux applications have been cross-compiled to run on the Green Board, adding most of the capabilities of a full-fledged desktop PC. Examples of applications that have been compiled and are available on each device in the network are SSH daemon, web server, network time protocol, GTK+, Video 4 Linux 2, gstreamer, ffmpeg, Open Sound System, tslib, Bluetooth, and various graphics and image processing libraries. An airplane-grade process monitoring system called monit is also utilized on the nodes, making sure everything is running properly.

The use of the Linux OS and a standardized development toolchain was essential in developing this system for lab-wide use. Many people with little to no knowledge of the underlying hardware have developed applications for this system, and multiple applications can be executing concurrently.

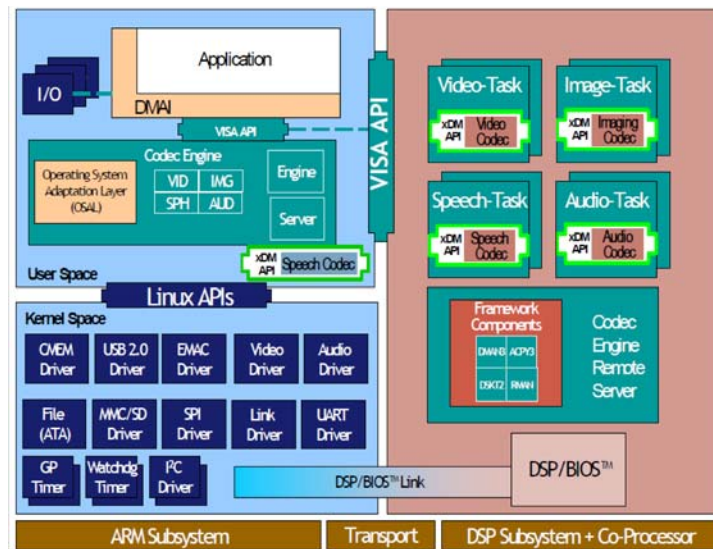
### 3.2.2.3 The Green Board - DSP/Video Code

Custom video devices can be very complicated to develop due to the large amount of time-sensitive data they need to handle. The SPINNER video network requires each video device to be able to capture audio and video from the microphones and camera, compress it, store it to an AVI file, packetize it into RTP packets, stream it using RTSP protocol, and capture still images into JPEG files... while SIMULTANEOUSLY decompressing, resizing, and

displaying a stream of audio and video from the network, looped back from the camera, or from a stored AVI file. In other words, it has to act like a video camera, still camera, audio recorder, IP-based webcam, and a networked media player all concurrently. These features also need to be abstracted and wrapped in a library so they are available for user applications.

The DSP core of the Davinci processor has the horsepower to execute the video processing functions required. The DSP core runs a small OS called DSP BIOS, which initializes the core and provides the basic functions of the chip to any code running on the DSP. An application called CODEC Server is the main system that runs on the DSP. CODEC Server maintains a registry of all the functions that have been coded and bootstrapped on to the DSP core. Client applications running on the ARM core can call these functions as if they were local, but they will run on the DSP core, which has access to the same memory as the ARM core, to pass data between the two.

FIGURE 3-13 Davinci Software Components Topology [77]





#### 3.2.2.4 Application Services

All of the capabilities of the device are made available to any user of the system via networked and local interfaces. A user of the system can access its features over the network through a standard sockets interface. For example, the user can tap into streams of sensor data, browse locally stored video clips, and snap JPEG stills by simply sending commands via TCP/IP without writing a single line of code. This type of service allows the system to be used as a facility by the entire community.

There are two main service applications. The first is called Spinner Data Server and it has the following responsibilities:

- 1 Control the Red Board - SPINNER Data Server's main task is to bridge the Red Board and the Green Board. This allows the sensor data and the channels of communication from the Red Board to be received and controlled by the Linux applications
- 2 Provide a sockets interface to features - SPINNER Data Server broadcasts real-time data from the Red Board and from any other sensors on a sockets server and receives control commands from a sockets server to control the appropriate feature
- 3 Register the device - SPINNER Data Server broadcasts periodic registration packets notifying the world that the device is available for use
- 4 Time synchronization - SPINNER Data Server communicates with a time broadcaster on the ethernet subnet using Network Time Protocol to monitor drift and network lag. It then broadcasts the exact time to any devices in wireless range
- 5 Control the motors and LEDs - SPINNER Data Server controls the actuators connected to the device. It performs any closed-loop control that has been enabled and receives socket commands for manual control

- 6 Connect to a database - If enabled, SPINNER Data Server can write directly to an SQL database any sensor data or event that is received, processed, or transmitted

The second service is called SPINNER Video Control and it is a very complex piece of code that performs the following functions:

- 1 Control the camera - SPINNER Video Control initializes the camera module and tells it when to snap a still frame and when to start capturing video frames. These frames are captured directly into contiguous RAM memory
- 2 Control the audio capture - SPINNER Video Control sets up the audio CODEC with sampling rate, channel count, gain, etc. and records blocks of audio into contiguous RAM memory
- 3 Control the speaker - SPINNER Video Control will stream raw audio samples from RAM memory to the audio CODEC for playback through a speaker or headphones
- 4 Control the video display - SPINNER Video Control can display RAW RGB video frames on the LCD
- 5 Control the resizer hardware - SPINNER Video Control can resize a RAW RGB video frame stored in RAM using the hardware resizer and put the resultant image back into RAM for display on the LCD
- 6 Compress video - SPINNER Video Control uses the algorithms in the DSP core to compress the video. Currently the DSP code has algorithms for MPEG-4 and H264 compression
- 7 Compress stills - SPINNER Video Control uses the algorithms in the DSP core to compress still images using JPEG compression and then writes the file to the FLASH memory

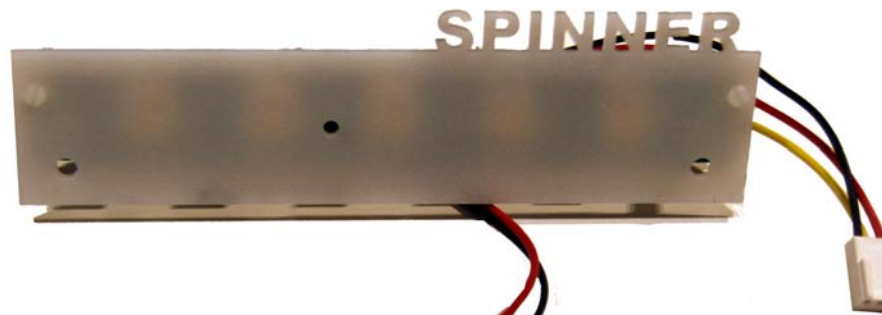
- 8 Compress audio - SPINNER Video Control uses the algorithms in the DSP core to compress the captured audio blocks using MP3, AAC, ADPCM, uLaw, or aLaw compression
- 9 Store compressed audio and video - After the video frames and audio blocks are compressed by the DSP code, the data is written to a properly formatted AVI file on the FLASH memory
- 10 Preview the video - SPINNER Video Control can display the captured frames in real-time on the LCD, useful for adjusting camera angles and testing the system
- 11 Stream the audio and video on the network - After the video frames and audio blocks are compressed by the DSP code, SPINNER Video Control wraps the data in RTP (Real-time protocol) packets with timing information and streams them on the network as a RTSP (Real-time streaming protocol) server. Audio and video served in this way can be received by standard players like VLC and Quicktime
- 12 Notify system when new data is available - SPINNER Video Control will announce to any attached clients when a new AVI file has been recorded or a new still has been captured. SPINNER Video Control can also write this information into any SQL database and push the files onto any server
- 13 Playback an AVI file - SPINNER Video Control can open an AVI file and read out frames of audio and video. It can decompress these frames using the DSP code, resize them using the resizer hardware, and display them on the LCD and speaker
- 14 Playback an RTSP stream - SPINNER Video Control can receive streaming audio and video from any RTSP server on the Internet, decompress the frames using the DSP, resize them, and then display them

- 15 Provide a simple sockets interface - SPINNER Video Control can be completely controlled over a sockets interface. This interface accepts high-level remote control commands such as snap JPEG, record video file, playback AVI file, and playback RTSP stream
- 16 Periodically execute commands - SPINNER Video Control can be told to periodically perform any of its functions. For example, it can be told to snap a JPEG every 30 seconds and start recording a 2-minute AVI file every 3 minutes

### 3.2.3 Modular Component #3 - The Hat

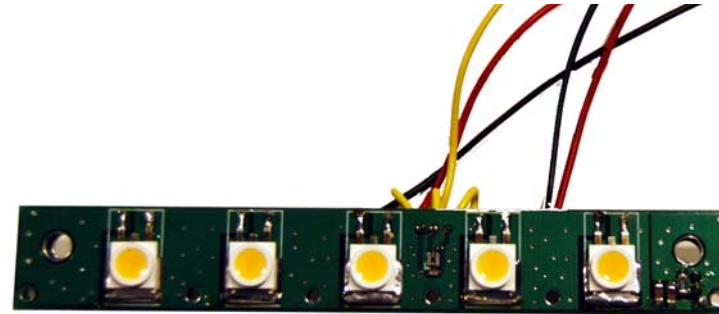
The third module that can be deployed as part of a distributed media system has two main functions, high-brightness illumination and the control of powerful DC motors. This module is known as the Hat since it normally sits on top of the video camera device.

FIGURE 3-14 Hat module with faceplate



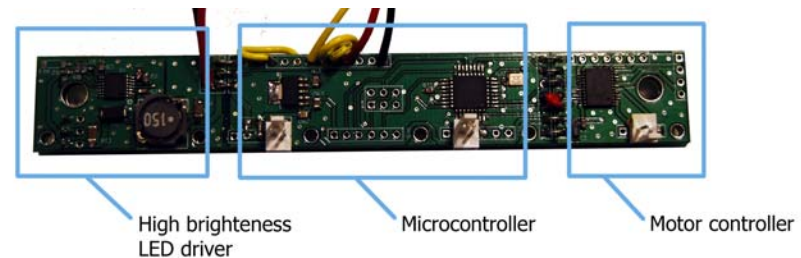
To provide illumination of the subject that is to be captured, the Hat contains five 1-Watt high-brightness LEDs and a LED switch-mode driver circuit. The driver circuit is dimmable by a PWM signal. There is also a photographic quality light sensor on the Hat should there be a need to provide closed-loop control of illumination based on ambient lighting conditions.

FIGURE 3-15 Front of Hat board with ganged high-brightness LEDs



The Hat board contains a small 8-bit AVR microcontroller. This microcontroller can send the appropriate PWM signals to the LED driver to control the brightness of the LED array. The microcontroller also monitors the temperature of the LED array and will turn it off should they get too hot. If set to full brightness for too long, the LED array, similar to a camera's flash bulb, can burn itself out or cause damage to nearby objects. The microcontroller can send a quick flash signal, fade the LED array up and down, pulse the LED array, or turn them completely on at a given brightness. The Hat board can be preprogrammed with screen savers or automatic behavior based on the light sensor, but it is normally controlled by another module such as the Green Board.

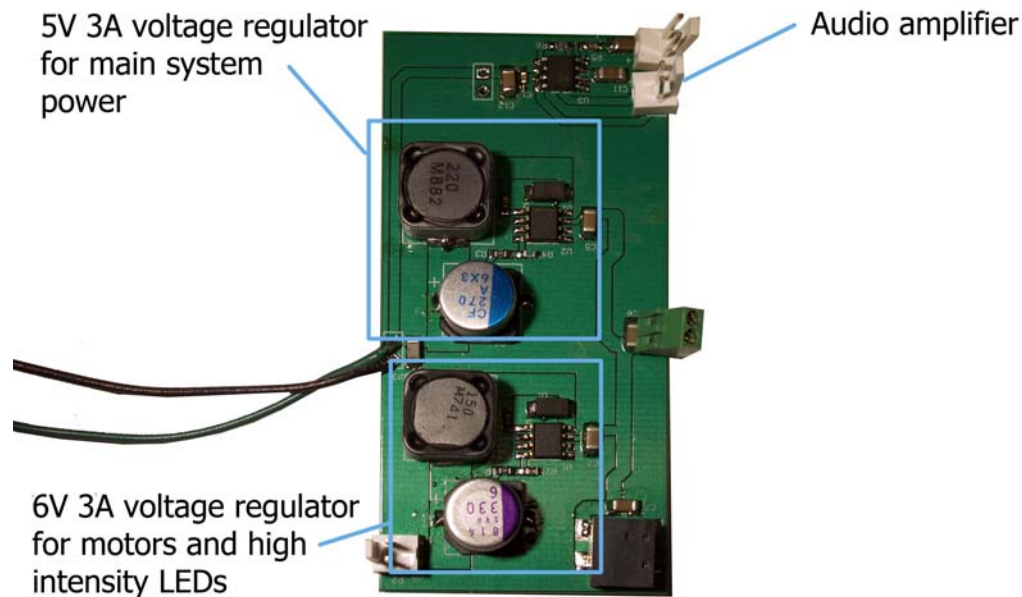
FIGURE 3-16 Back of Hat board with electronics



In addition to the illumination system, the Hat also can control two DC motors. The Hat has an onboard DC motor driver, which can control the speed and direction of the two motors. This motor driver circuit is connected to the Hat's microcontroller. The Hat has connections for several types of position sensors, including rotary encoders and potentiometers. The Hat microcontroller uses PID motor control code to operate the motor as a servo. The Hat's motor controller can drive high-torque geared motors for the panning and tilting of heavy cameras. The motors can be controlled by an external system by sending a position in 1/10th degrees to the Hat's microcontroller.

### 3.2.4 Modular Component #4 - The Power Supply

FIGURE 3-17 Power Supply Board Diagram



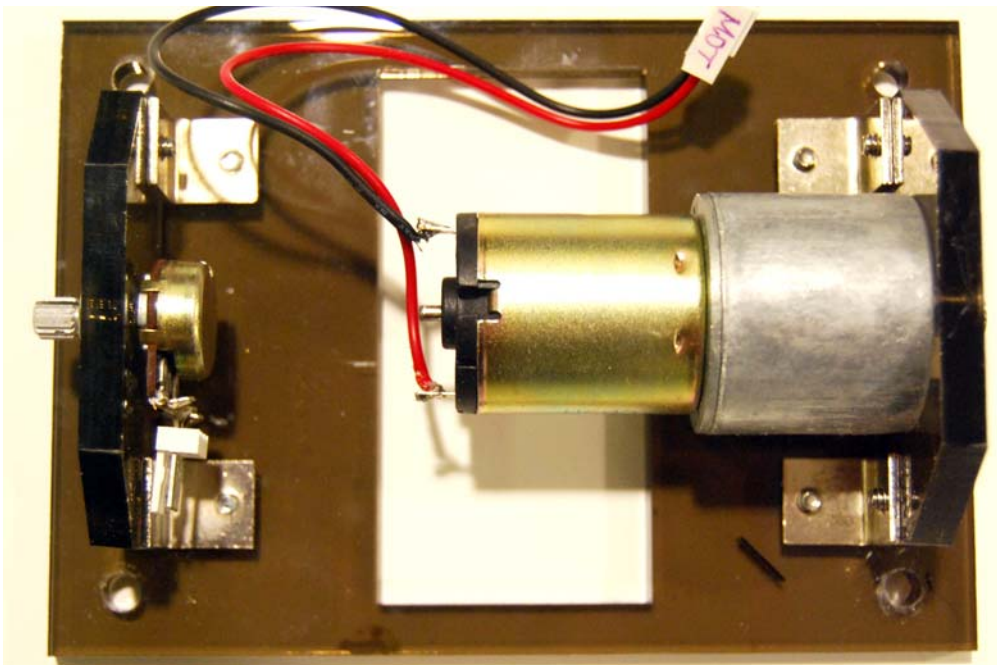
To power all the modules and support the addition of mechanical components such as motors for panning and tilting, a separate power supply board has been designed. This board takes an input voltage from 8VDC to 40VDC, allowing it to be driven by many DC sources including Power-Over-Ethernet. The input voltage is bucked down into two rails at 5V and 6V output voltages supporting up to 3 Amps each. The Power Supply board also has an audio amplifier to drive a speaker. The audio output of the Green Board is amplified and sent to a speaker on the faceplate of the ubiquitous media portal.

### 3.2.5 Complete Portal Assembly

A Ubiquitous Media Portal unit is created by assembling the above described modules with the following mechanical components:

- 1 Wall mount plate - The wall mount is a custom fabricated plate that attaches to the wall, ceiling, furniture, or any surface with four screws or with ultra-high bond double stick tape. The wall mount contains a gearhead motor that is geared to spin at 6 RPM with enough torque to move 24kg. This motor is used to provide a tilt feature to the portal's video camera by tilting the entire unit. This motor can also be used to "slump" the camera down indicating that it is turned off for privacy. The wall mount bracket also has a potentiometer in axis with the motor for rotational position indication. The wall mount has a cutout for the Power Supply board to be mounted on the wall with the bracket

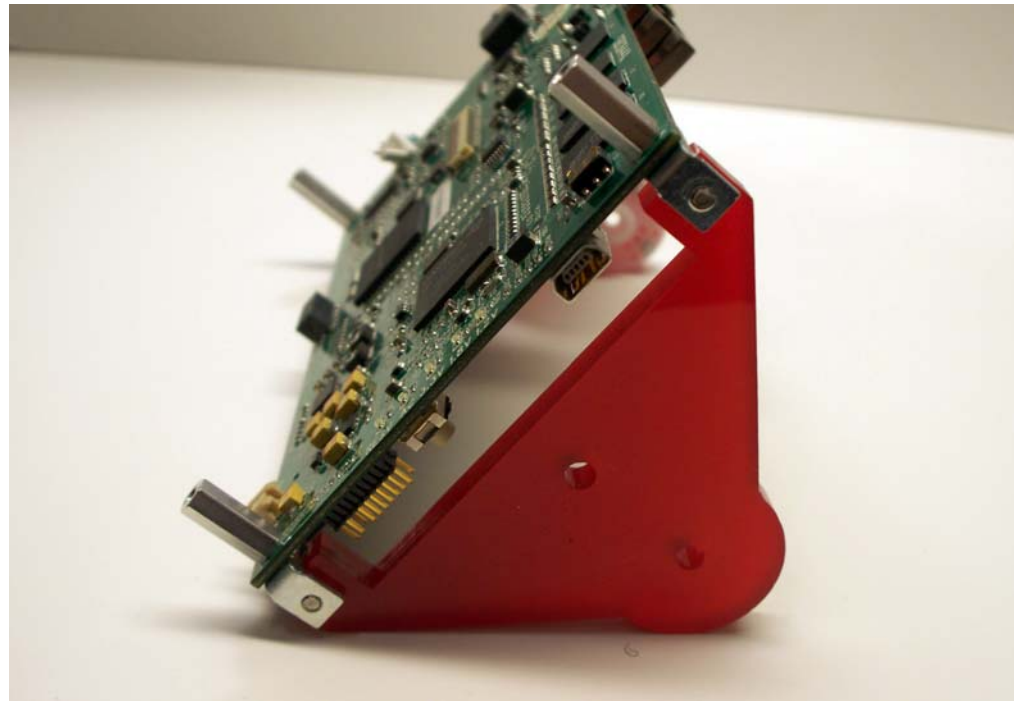
FIGURE 3-18 Wall Mount



- 2 Side pieces - The Green Board is bolted to two side pieces. These side pieces attach to the shaft of the motor and the shaft of the potentiometer on the wall mount. Two set screws finalize the attachment. This system forms the mechanism for the tilting action of the portal and is how the portal is mounted on the wall. The portal can quickly be removed from the wall mount should a need arise



FIGURE 3-19 Green Board attached to side pieces



- 3 Front plate - The LCD touchscreen, a speaker, and a small servo motor attach to a custom fabricated plastic plate. The camera module mounts to the servo on this front plate so that the camera can be panned. The front plate also provides routing for the cameras cable to the Green Board. This complete front plate assembly is bolted with the Red Board to the Green Board/side piece assembly

FIGURE 3-20 Front plate with attached speaker, LCD, and panning camera

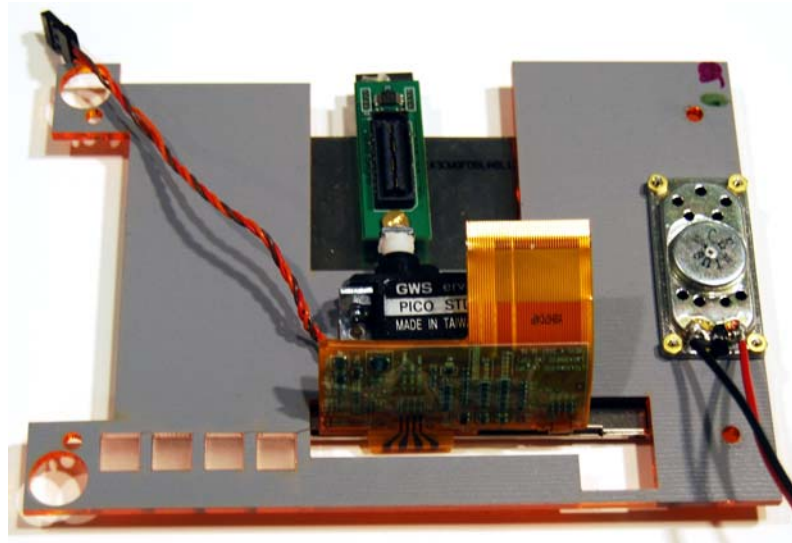


FIGURE 3-21 Front plate with Red Board ready to assemble



- 4 Metal armor - The complete assembly is covered with a set of custom fabricated and patined metal pieces. This cover protects the circuitry, hides the wires, and provides the design aesthetic to help with the system's acceptance. The Hat module is attached to the top metal piece with two corner brackets

**FIGURE 3-22** Side view of complete mounted assembly with armor

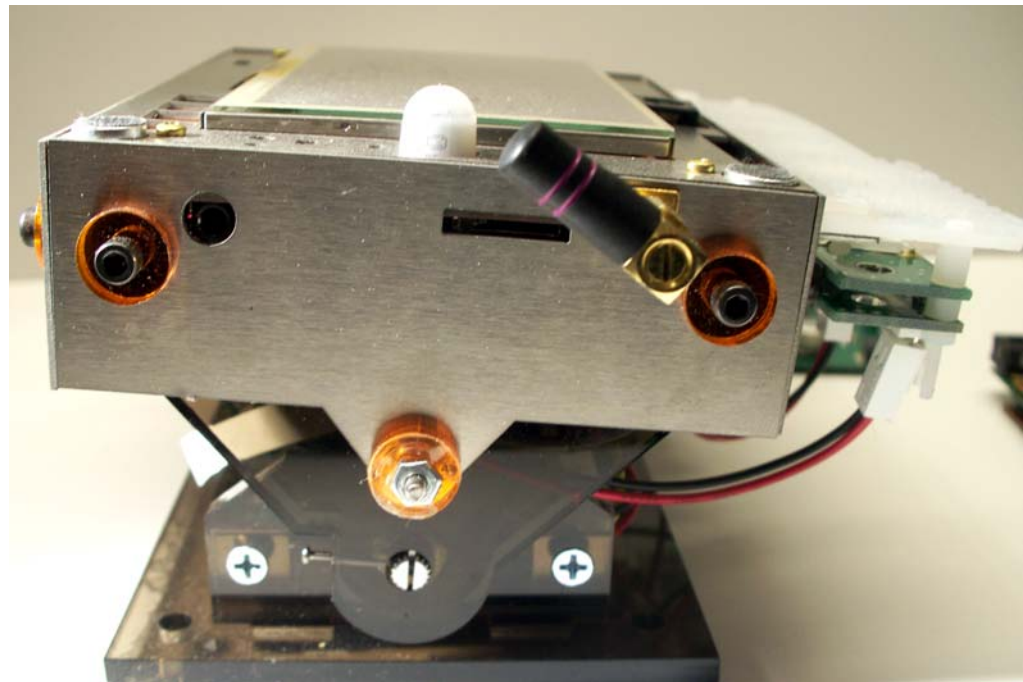


FIGURE 3-24 Completed Ubiquitous Media Portal



FIGURE 3-23 Portal in action, illuminating subject for video capture



Once assembled, 45 units were deployed around the MIT Media Lab buildings covering all populated areas. The building's system administrators provided the SPINNER network with its own dedicated subnet.

## 3.3 WEARABLE SENSOR DEVICES

To capture the data most pertinent to the detection of human behavioral events and context, a system of wearable sensors was developed. These on-body sensing devices provide direct access to the subjects and can identify a participant to the system with location information.

The combination of wearable sensing with a distributed sensor network situated in the environment is one of the major achievements of this research. Besides capturing the behavioral data only accessible with wearable sensors, the on-body devices augment the overall network by recording subjective media such as audio directly from the participant that can be synchronized with media captured throughout the system.

To support a wide range of applications in wearable sensing, two devices were developed for this research program. These devices, detailed below, are called the SPINNER microBadge, or  $\mu$ Badge, and the SPINNER wrist sensor unit.

### 3.3.1 The $\mu$ Badge

The first wireless wearable sensor device developed is in the form of a small identification badge. This badge is attached to the torso of the wearer in one of three ways:

- 1 Lanyard - can be hung around the neck using string, chain, or a lanyard
- 2 Pin - can be pinned to wearer's shirt or jacket
- 3 Pocket - can be placed in a shirt pocket provided the material is not opaque to infrared, most shirts work without a problem

FIGURE 3-25 SPINNER Wearable sensors





The  $\mu$ Badge builds upon existing badge platforms to investigate conversational dynamics and social signaling [52] [53]. The basic features of the  $\mu$ Badge are:

- 1 a suite of motion and orientation sensors
- 2 integration with the SPINNER media network through a ZMat RF transceiver
- 3 high quality audio capture of wearer's speech
- 4 onboard storage can store up to two weeks of audio and data
- 5 dedicated audio CODEC and DSP for capturing and analyzing speech in real-time
- 6 IR transceiver for line-of-sight communication
- 7 user-input switches
- 8 speaker
- 9 OLED display

FIGURE 3-26  $\mu$ Badge pinned to participant



FIGURE 3-27 Diagram of the microBadge (front)

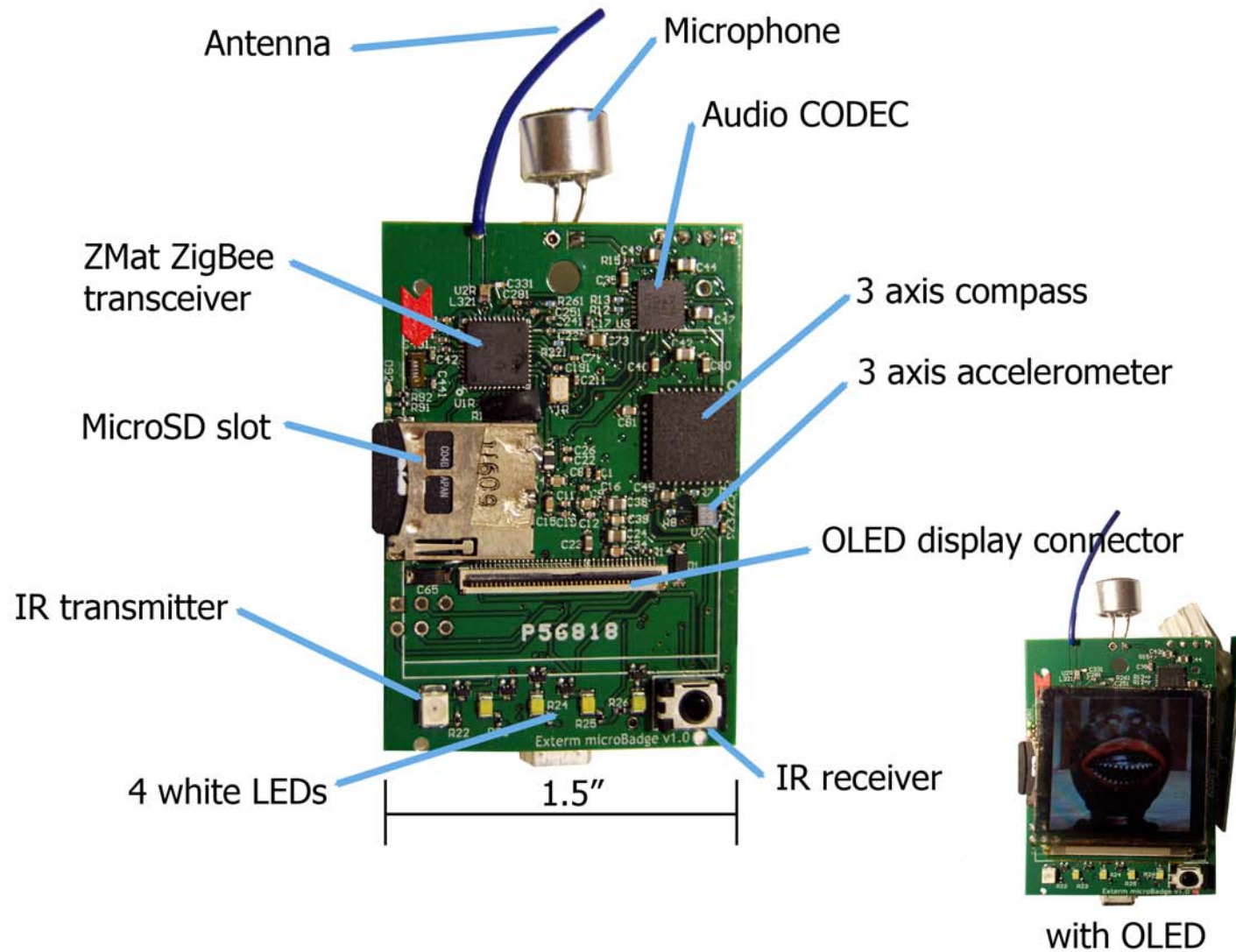
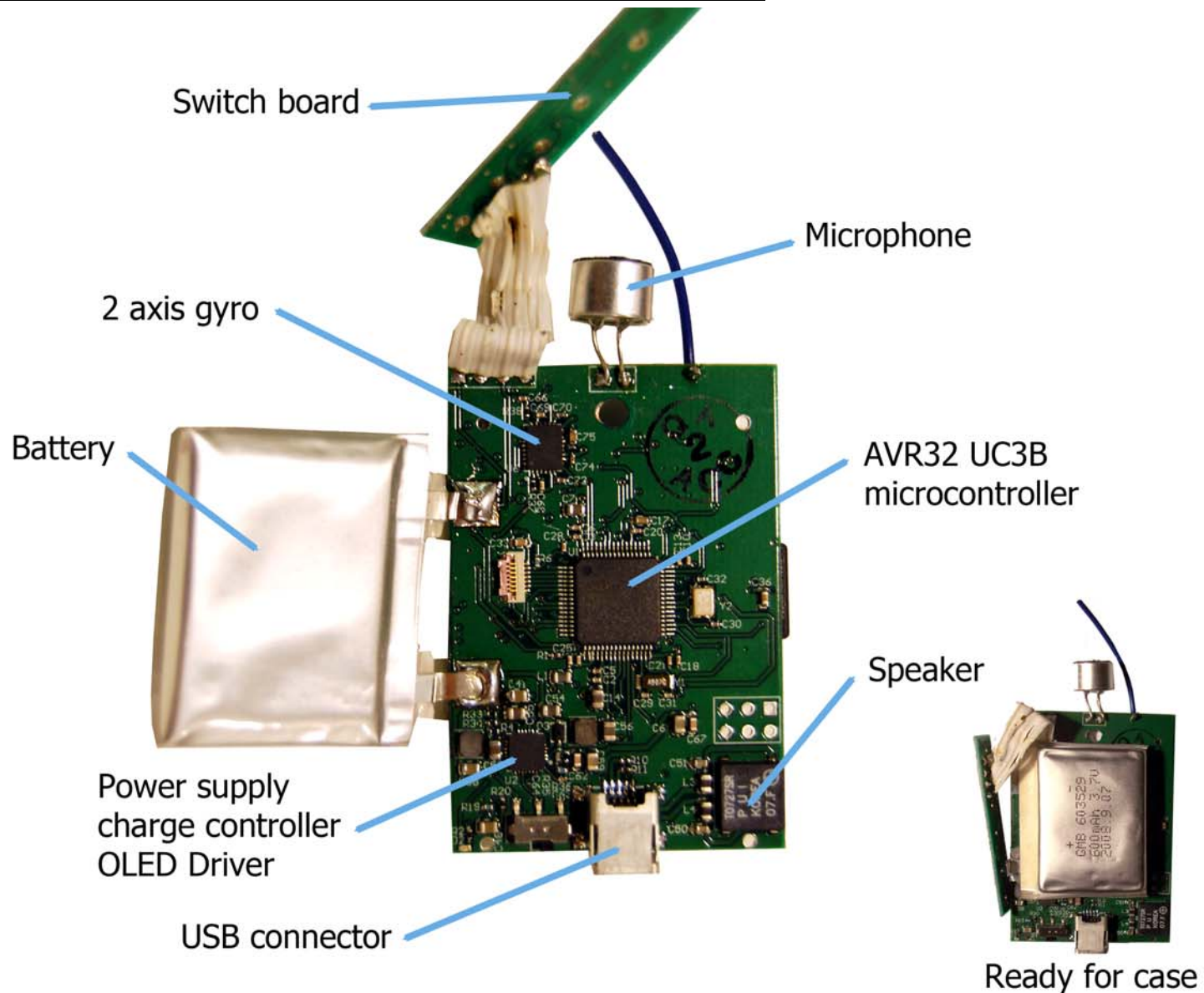


FIGURE 3-28 Diagram of the microBadge (back)





### 3.3.1.1 microBadge Sensor Suite

The main task of the  $\mu$ Badge is to sample its onboard sensors, store the captured sensor data to its onboard flash, analyze the data if real-time features are required, and transmit any resultant information when necessary. The  $\mu$ Badge has the following sensors:

- 1 Gyroscope, 2-axis angular velocity sensor. The IDG-500 factory calibrated gyro from Invensense can be read by the  $\mu$ Badge microcontroller as fast as 10000 samples per second with 16-bit resolution. The default sample rate is set to 100 samples per second which has proven adequate considering the mounting position of the badge on the torso
- 2 Accelerometer, 3-axis acceleration sensor. The VTI CMA3000 accelerometer is the world's smallest accelerometer. It has a digital interface and a maximum sampling rate of 400 samples per second at 8-bit resolution. The default sampling rate of this sensor is also set at 100 samples per second which more than accurately captures vibration, jitter, and the relatively slow body motion of the torso
- 3 Compass, 3-axis global orientation sensor. The Honeywell HMC3643 provides 10 samples per second of world-view 3-axis orientation. The wearer's heading is easily read from this sensor which can be used across multiple  $\mu$ Badges to capture the relative orientation between multiple participants engaged in a conversation or any group social activity
- 4 Infrared transceiver. The  $\mu$ Badge uses an IR emitter and receiver to identify itself to whatever the wearer is directly interacting with, either a SPINNER media portal or one or more  $\mu$ Badges worn by other people. Therefore each device also knows what or who is directly interacting with it. This system can be used to identify when a conversation is taking place, who is involved, and the duration of the encounter

### 3.3.1.2 microBadge Audio System

The audio system of the  $\mu$ Badge acts as both a sensor device and a media capture system. A microphone is positioned on the top edge of the badge pointing directly up at the badge wearer's face. The directional microphone captures the wearer's speech with minimal interference from ambient sounds or other people conversing with the badge wearer. When multiple people are wearing badges in a conversation, the audio can be isolated to each participant or be resynched using timing information shared wirelessly amongst the badges.

The microphone is attached to a dedicated audio CODEC chip, the Texas Instruments PCM3793, which provides an automatic gain controlled microphone amplifier, a notch filter to remove noise, and a CD quality analog-to-digital converter. The digital audio is then sent through a dedicated I2S interface to an AVR32 UC3B microcontroller.

The microcontroller has a built-in set of DSP instruments that can perform FFTs, band-specific RMS power calculations, and look for changes in spectral characteristics of the speech. After any real-time processing is completed, the DSP section of the AVR32 is used to compress the audio with a standard ADPCM compression algorithm. The compressed audio is stored to the microSD card along with the sensor data.

The badge samples the audio at 12096 samples per second at 16-bit resolution. The ADPCM algorithm applies a 4-to-1 compression ratio resulting in 6048 bytes of storage used per second. The badge can support up to 16GB of Flash storage which can store around one month of constant audio capture.

The audio system can also be used to play back audio as a user interface feature to signal the arrival of information, a change in state, or to audition any captured audio.

### 3.3.1.3 microBadge ZMat Wireless Transceiver

The badge contains the same ZMat wireless module as the Red Board described in section 3.2.1.4. The ZMat system on the wearable devices, unlike on the Red Board, is programmed as a ZigBee end-device.

As an end-device, the wireless transceiver operates in a low-powered mode with the radio defaulting to sleep mode until the device needs to send a packet. To receive a message, the wearable devices send out a polling command to any nearby ZigBee routers in the mesh network checking if there is any data addressed to the end-device. After sending a polling command, the end-device keeps the radio enabled in receive mode for a few milliseconds to wait for any messages.

FIGURE 3-29  $\mu$ Badge in its case, close-up



The main function of the ZMat wireless transceiver on a wearable device is to integrate the device into the SPINNER network. It does this by sending out periodic pings to identify the device to the network and allow the location system to track the mobile wearable sensor. If a ping is received by more than one basestation, the basestations send each other the signal strength of the ping using their wired network. The location of the device is approximated by triangulating the signal strength data amongst the basestations that received the ping. The accuracy of this system is dependent on the density of basestations to increase the number of nodes that can receive a ping and the amount of signal strength calibration data available for the environment. In the more densely covered areas of the deployed SPINNER network, the system has repeatedly shown 3 meter accuracy with unstructured “naked eye” testing. Rigorous testing of the localization system in real multipath-laden and dynamic environments has shown much less accuracy, e.g. 4 meters. Similar systems with proper calibration and estimates have shown accuracy better than 3 meters [78]. The default ping interval is 10 seconds, but can be changed according to a specific application’s average speed of motion of the wearable devices in the environment.

To complete the integration with the SPINNER media network, the wearable devices need to be time synchronized with the system. In addition to the identification and location pings, the ZMat transceiver in the end-device sends out periodic requests for the current time. Any Red Board that receives this request will respond with the current time with millisecond accuracy. The Red Boards are all synchronized using Network Time Protocol over their wired ethernet. Upon receiving the time, the wearable device sets its clock and then keeps its own time with an accurate crystal-based real-time clock. Network delays are approximated and removed by comparing the interval between two clock set messages with the real-time clock's interval.

In addition to this integration with the larger system, the ZMat transceiver can be used to off-load recorded data, transmit real-time information, stream audio, and update the device's firmware.

#### **3.3.1.4 microBadge Additional Features**

The  $\mu$ Badge contains several other features required for its operation and to support additional applications.

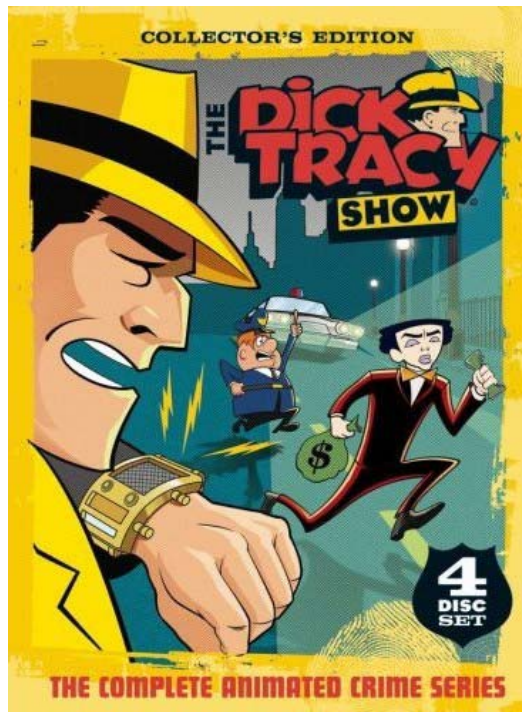
- 1 Power Supply - The  $\mu$ Badge's power supply uses a 600mAh Lithium polymer rechargeable battery. It contains a switchmode regulator providing the system voltage of 3.3V and a boost regulator providing the 17V required to drive the OLED display. It also has an onboard battery charge controller and battery protection circuitry
- 2 OLED Display - The  $\mu$ Badge sports a 1.5 inch diagonal 128x128 pixel 16-bit color RGB OLED display
- 3 LEDs - 4 status LEDs are included and can display animations such as an audio level indicator
- 4 Switches - 3 switches can be used for user input

- 5 USB - A USB connection can be used to charge the battery, read the onboard Flash memory, upload new firmware, and integrate the device with a larger system

### 3.3.2 Wrist Sensor Unit

The second wearable device developed for this research is the wrist sensor unit. It is intended to add to the system gestural and biometric sensing channels that can not be seen from the torso mounted badge. By positioning additional sensors on the wrist, the system can now react to manipulator signaling, specific gestures, context indicated by arm activity, and galvanic skin response.

**FIGURE 3-30** The wrist sensor unit has a speaker, microphone, display, buttons, and wireless communication, making it quite similar to the fictional Dick Tracy wrist communication device [79]



**FIGURE 3-31** Wrist Sensor Unit

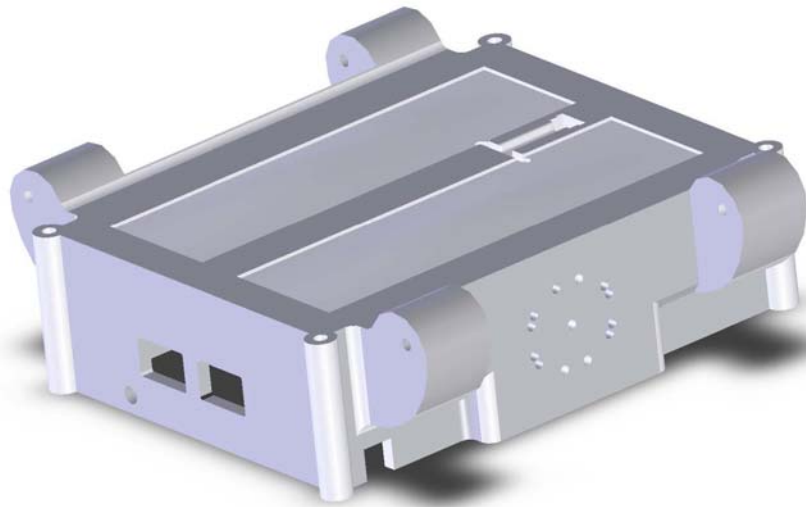


The wrist sensor unit is based on the  $\mu$ Badge and contains the exact same ZMat transceiver module to integrate it with the SPINNER network. It uses the same microcontroller and has the same additional features such as a microSD card slot, USB port, power supply, switches, and LEDs. Since it is in a slightly larger package and oriented in a landscape manner on the wrist, the OLED display is slightly larger with 160x128 pixels.

Although the wrist sensor unit has the same audio CODEC as the  $\mu$ Badge, it does not have a microphone in the standard configuration and does not normally capture audio, leaving this feature to the  $\mu$ Badge which is better positioned closer to the source. It has an onboard speaker which can be used in conjunction with the OLED display to play back audio and video captured by the system.

A custom case was designed and fabricated using the Dimension Systems 3D printer. Using injection molding the case can be streamlined to the size of a normal wrist watch.

**FIGURE 3-33 CAD rendering of case bottom prior to printing**



**FIGURE 3-32 CAD rendering of the case top**

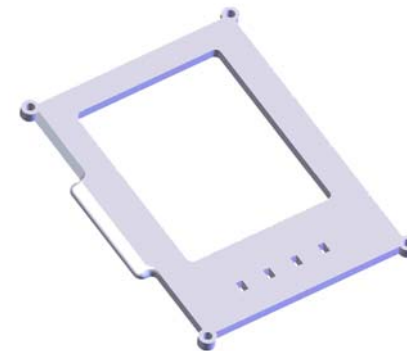




FIGURE 3-34 Diagram of wrist sensor unit (front)

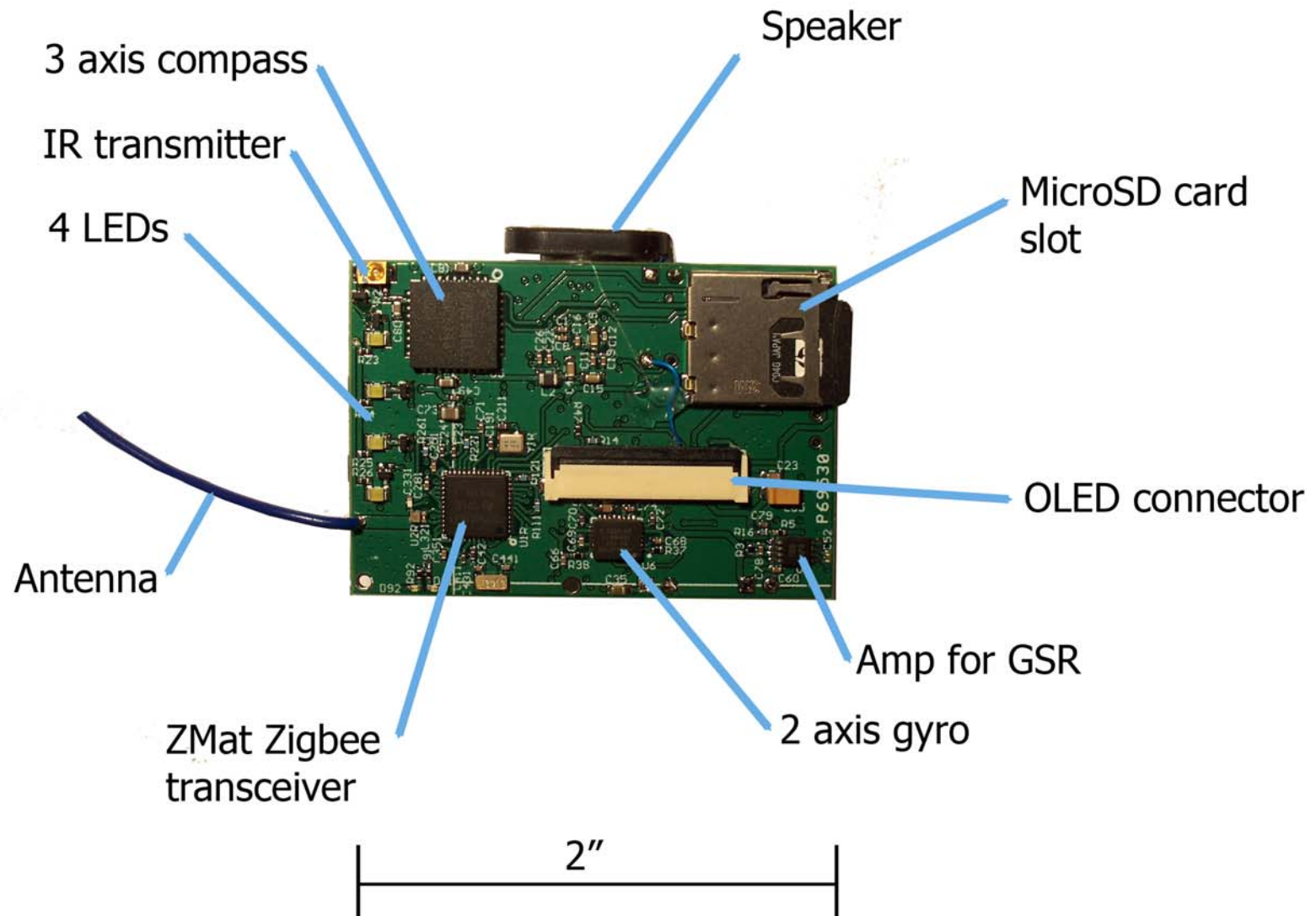
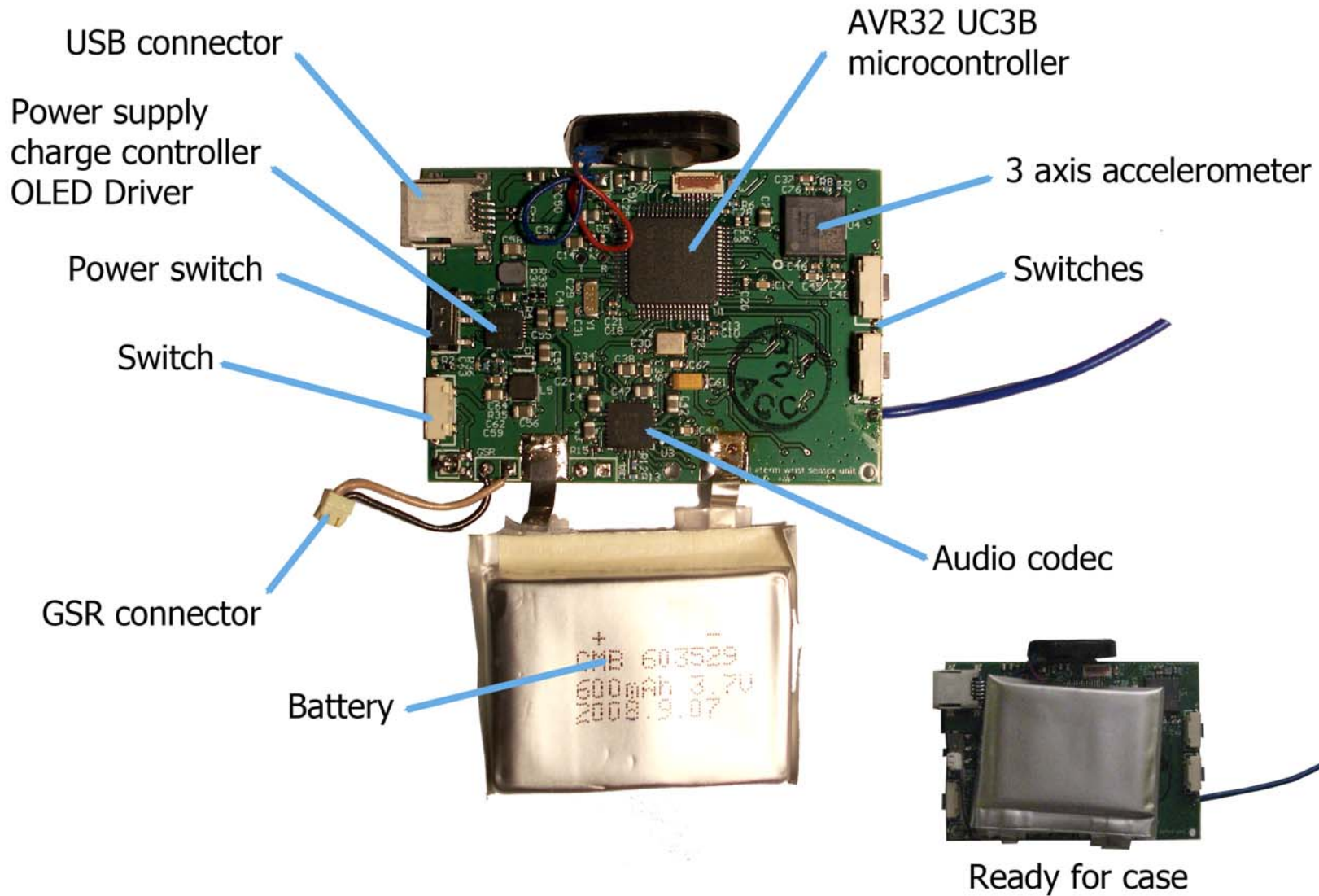


FIGURE 3-35 Diagram of wrist sensor unit (back)





### 3.3.2.1 Wrist Unit Sensor Suite

The sensors used by the wrist unit vary somewhat from the  $\mu$ Badge. The sensors on the wrist unit are:

- 1 Galvanic Skin Response, The wrist unit has GSR sensing built directly into the watch band. It does not require any sticky probes and it is mostly unnoticeable to the wearer. The GSR sensor samples skin conductivity at 50 samples per second and can indicate a change in physiological state often related to stress, internal struggle, or excitement/arousal

FIGURE 3-36 Wrist unit's GSR probes on arm side of device



- 2 Accelerometer, 3-axis acceleration sensor. The VTI SCA3100 is a high-end factory calibrated accelerometer. It has a digital interface and a maximum sampling rate of 1000 samples per second at 12-bit

resolution. The default sampling rate of this sensor is set at 500 samples per second, which more than accurately captures hand motion, gestures, and hand shakiness

- 3 Gyroscope. 2-axis angular velocity sensor. The IDG-500 factory calibrated gyro from Invensense can be read by the  $\mu$ Badge microcontroller as fast as 10000 samples per second with 16-bit resolution. The default sample rate is set to 1000 samples per second which is more than adequate for human arm motion
- 4 Compass. 3-axis global orientation sensor. The Honeywell HMC3643 provides 10 samples per second of world-view 3-axis orientation. This can be used in the wrist watch to provide a world orientation to base the inertial sensors as well as additional gesture information
- 5 Infrared transmitter. The wrist unit has an IR transmitter to send line-of-sight messages to other devices. Unlike the  $\mu$ Badge it does not have an IR receiver as its visibility is often unknown and obscured.

### 3.3.3 Battery Life

Both the wrist sensor unit and the  $\mu$ Badge use a 600mAh lithium polymer battery. The radio runs in a low-power end device mode which remains off in between transmissions and periodic polls for incoming packets. With the OLED display turned off and all the sensors on at their default sampling rate as described previously, the average current draw for both the  $\mu$ Badge and the wrist sensor unit is approximately 50mA. Testing of the devices have shown that in this state they will run and collect data continuously for greater than twelve hours on a fully charged battery. The battery can be recharged using USB, taking 1.5 hours to fully recharge an empty cell.

## 3.4 BACKEND SERVICES

The devices designed for this system come in a variety of configurations falling into two major tiers of service depending on function and available resources. As described above, one tier contains nodes connected to the local LAN with permanent power providing situated and higher-bandwidth media services. The second tier contains mobile, wearable nodes communicating via wireless powered by small batteries providing on-body access to personalized human behavioral information. All of the devices in these two tiers have enough processing capability to analyze in real-time the data they capture or analyze the stored data they have collected already. This allows the application on the local node to dynamically change its behavior and minimize required communication by filtering data and processing locally. In any distributed system, applications exist that require an aggregate of data from multiple, if not all, nodes in the system. While this aggregation can happen within the distributed network, many applications, including the SPINNER video application, will require additional processing and massive storage to be performed with an omniscient view of the system.

Therefore a third tier of services was added to the system. This tier provides centralized services, scalable storage and processing capabilities, the use of standard data services, and a single point of access to the entire system for use by end-user applications and by the distributed nodes in the system.

The server needs to be powerful enough to enable processing and storage features that cannot easily be done within the network on the embedded devices. The server used in the SPINNER system has 8 CPUs, 16GB of RAM, dual gigabit ethernet ports, and a 5TB RAID array which can easily be expanded. Since the SPINNER system captures and stores audio and video, there is a major storage requirement. The server runs Red Hat Enterprise

Linux, which is stable, fast, and supports endless open source and commercial applications.

### 3.4.1 Relational Database

The centralized computer runs a MySQL relational database server. This database is a robust, high-performance solution for aggregating, warehousing, and serving the data and multimedia from the entire distributed system.

MySQL has adapters in every language and for every platform that allow devices and applications to communicate with the database directly, in parallel, and at speeds up to the entire network bandwidth. Every device in the SPINNER system, whether a battery-powered wearable or a video camera device, can write data directly to the database from anywhere in the network.

MySQL is an industry standard database supported by endless development tools and software packages. This makes it easy for anyone with little to no knowledge of the complexities of the distributed sensor and video network to access the data for their applications and visualizations. The database can be written to and queried fast enough to support real-time applications using the database as middleware.

Besides being populated with raw sensor data, network events, and media files, the database can be updated with the results of analysis tools. These tools are custom pieces of software that query the database for new data, map the data to higher-level information, or otherwise classify the data. The results of these algorithms and programs can then be stored in the database along with the raw input and the video files. This creates an ideal setup for testing sensor mapping algorithms as you are left with an easy to view collection of input, output, and a video record of the analyzed event.

FIGURE 3-37 SPINNER Database Schema Diagram

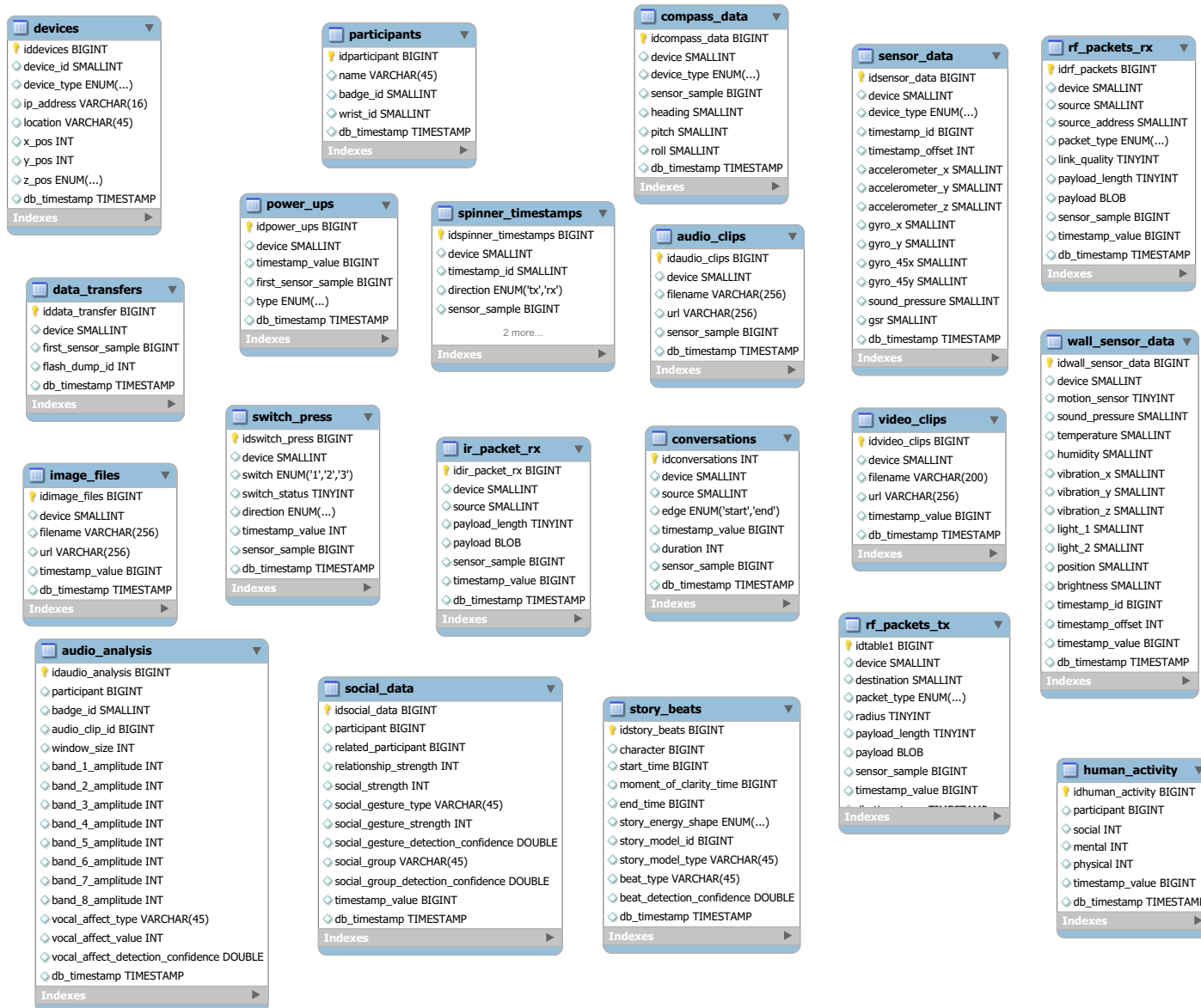


FIGURE 3-38 SPINNER Database MySQL sensor data query view

MySQL Query Browser - Connection: simonetti / spinner\_db

File Edit View Query Script Tools Window Help

Transaction Explain Compare

Resultset 1

SQL Query Area

```
SELECT * FROM sensor_data s LIMIT 1000000;
```

id	sensor_data	device	device_type	timestamp_id	timestamp_offset	accelerometer_x	accelerometer_y	accelerometer_z	gyro_x	gyro_y	gyro_45x	gyro_45y	sound_pressure	gsr	db_timestamp
158922	40631	wrist	3316658	1999	-256	352	-10464	396	415	335	432	0	931	2009-11-22 23:21:35	
158923	40631	wrist	3316658	0	-152	352	-10464	396	415	332	432	0	931	2009-11-22 23:21:35	
158924	40631	wrist	3316658	1	-200	312	-10496	396	415	333	433	0	932	2009-11-22 23:21:35	
158925	40631	wrist	3316658	2	-304	224	-10424	395	415	332	432	0	931	2009-11-22 23:21:35	
158926	40631	wrist	3316658	3	-280	264	-10384	395	415	331	428	0	934	2009-11-22 23:21:35	
158927	40631	wrist	3316658	4	-328	200	-10384	396	415	334	432	0	931	2009-11-22 23:21:35	
158928	40631	wrist	3316658	5	-232	256	-10488	396	415	332	428	0	931	2009-11-22 23:21:35	
158929	40631	wrist	3316658	6	-320	296	-10400	396	415	332	432	0	935	2009-11-22 23:21:35	
158930	40631	wrist	3316658	7	-368	312	-10424	396	415	332	432	0	931	2009-11-22 23:21:35	
158931	40631	wrist	3316658	8	-216	336	-10432	396	415	333	432	0	931	2009-11-22 23:21:35	
158932	40631	wrist	3316658	9	-176	368	-10496	396	415	334	432	0	931	2009-11-22 23:21:35	
158933	40631	wrist	3316658	10	-176	264	-10424	396	415	332	432	0	931	2009-11-22 23:21:35	
158934	40631	wrist	3316658	11	-280	288	-10384	396	415	332	432	0	933	2009-11-22 23:21:35	
158935	40631	wrist	3316658	12	-336	240	-10400	396	415	334	432	0	931	2009-11-22 23:21:35	
158936	40631	wrist	3316658	13	-288	248	-10448	396	415	332	430	0	931	2009-11-22 23:21:35	
158937	40631	wrist	3316658	14	-296	248	-10424	396	415	332	432	0	933	2009-11-22 23:21:35	
158938	40631	wrist	3316658	15	-288	232	-10424	396	415	332	428	0	931	2009-11-22 23:21:35	
158939	40631	wrist	3316658	16	-264	344	-10424	396	415	332	430	0	934	2009-11-22 23:21:35	
158940	40631	wrist	3316658	17	-360	312	-10384	396	415	335	429	0	931	2009-11-22 23:21:35	
158941	40631	wrist	3316658	18	-224	344	-10456	396	415	332	429	0	931	2009-11-22 23:21:35	
158942	40631	wrist	3316658	19	-152	312	-10512	396	415	332	432	0	934	2009-11-22 23:21:35	
158943	40631	wrist	3316658	20	-184	352	-10456	396	415	332	432	0	931	2009-11-22 23:21:35	
158944	40631	wrist	3316658	21	-232	336	-10464	396	415	332	430	0	931	2009-11-22 23:21:35	
158945	40631	wrist	3316658	22	-256	312	-10576	396	415	334	432	0	931	2009-11-22 23:21:35	
158946	40631	wrist	3316658	23	-248	264	-10400	396	415	332	430	0	932	2009-11-22 23:21:35	
158947	40631	wrist	3316658	24	-216	216	-10392	396	415	332	432	0	933	2009-11-22 23:21:35	

1000000 rows fetched in 8.398751s (0.0030s)

1: 1

Edit Apply Changes Discard Changes First Last Search

Schemata

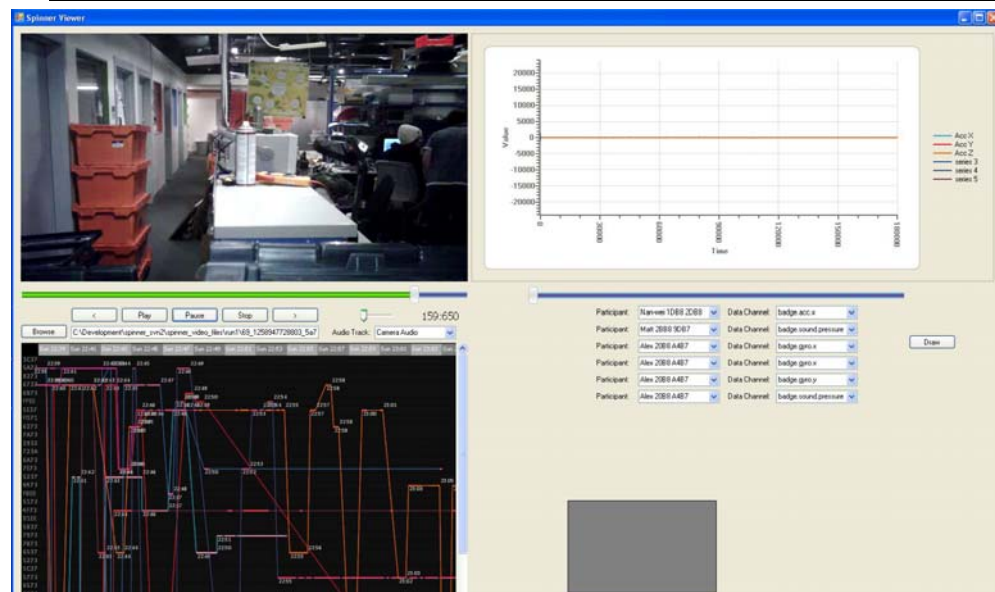
- information\_schema
- mydb
- mysql
- spinner
- spinner\_bit
- spinner\_data
- spinner\_db
  - audio\_analysis
  - audio\_clips
  - character\_location
  - compass\_data
  - conversations
  - data\_transfers
  - devices
  - human\_activity
  - image\_files
  - ir\_packet\_tx
  - participants
  - personal\_data
  - power\_ups
  - rf\_packets\_tx
  - rf\_packets\_tx
  - sensor\_data
  - social\_data
  - spinner\_timestamps
  - story\_beats
  - switch\_press
  - video\_clips
  - wall\_sensor\_data
  - spinner\_wearable\_data

Syntax Functions Params Trx

- Data Definition Statements
- Data Manipulation Statements
- MySQL Utility Statements
- MySQL Transactional and Locking ...
- Database Administration Statements
- Replication Statements
- SQL Syntax for Prepared Statements



**FIGURE 3-39 SPINNER Data/Video Viewer**



### 3.4.2 SPINNER Daemon

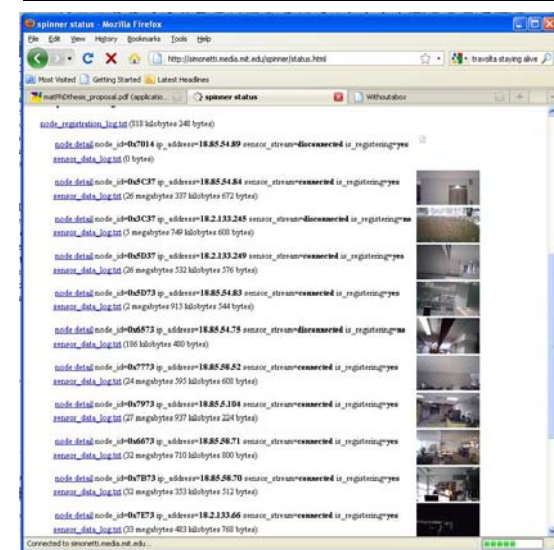
The SPINNER Daemon software, or `spinnerd` for short, implemented by fellow Responsive Environments Group RA, Bo Morgan, provides status information about the entire system and acts as a gateway for applications to quickly interact with the system.

Spinnerd maintains a registry of available devices in the distributed sensor network. It receives periodic pings and startup announcements from the nodes in the system. When it discovers a sensor device such as a Red Board, `spinnerd` queries its capabilities and binds a socket connection. The device will then stream to `spinnerd` all collected real-time sensor data and any packets received from wireless nodes in its vicinity. Spinnerd also receives notifications from devices that have a newly captured video clip, audio file, or still image. Upon receiving these notifications, `spinnerd` will transfer the new file and store it on the server.

Spinnerd makes all of the data, packets, and files received from the entire system available via a browsable web interface, a file transfer protocol, and a custom sockets interface directly to `spinnerd`. The web interface is an essential tool for maintaining the system, as it gives up-to-the-second operational status information on all nodes. The sockets interface can be used by any device or software application that quickly needs system-wide information. Examples of the type of system-wide information that `spinnerd` serves in real-time are which mobile devices are within camera/interaction range of which situated nodes, and the amount of activity (a combination of motion sensor readings and audio dynamics) witnessed by every node in the system.

In addition to being a source of “omniscient” information, `spinnerd` can also send commands to one node, multiple nodes, or all nodes in the network. These commands can move the pan/tilt motors, turn on/off sensors, completely disable a node, display information on a remote screen, control the lighting, and activate any feature that the device supports. Although these commands

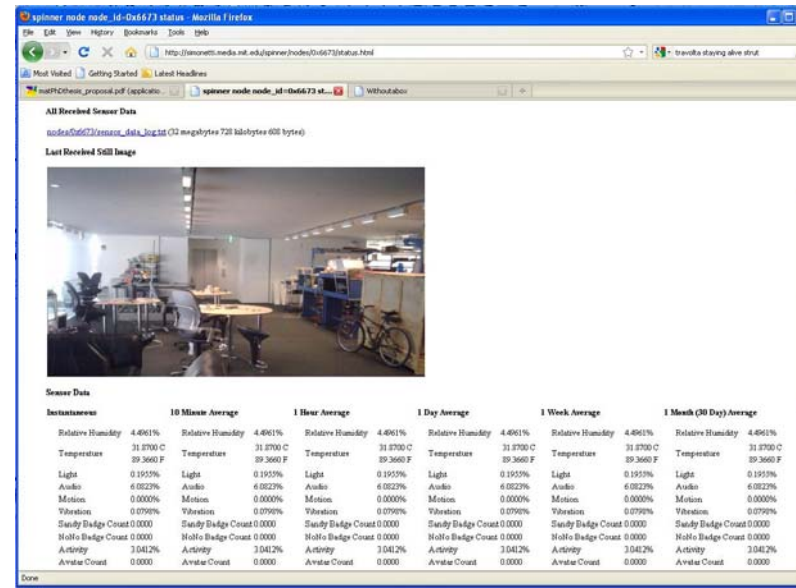
FIGURE 3-40 Spinnerd status page





can also be sent directly to the device itself, spinnerd provides a single point of connection for a client application and provides a registry of devices, sensor data, system status, and a quick way to send commands to registered devices.

FIGURE 3-41 Spinnerd node report



## 3.5 PLATFORM APPLICATIONS

---

During the design, construction, and deployment of the SPINNER media network, many applications and demos were developed to test the operation of the platform and aid in its acceptance by providing value-added services. These applications are summarized below, with the main SPINNER application being detailed in Chapters 4, 5, and 6.

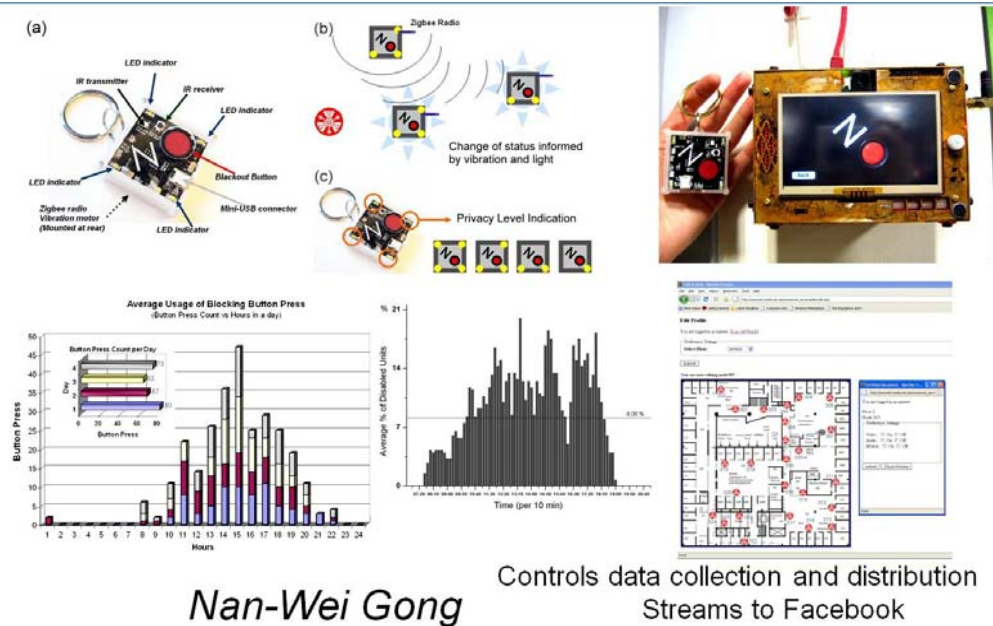
### 3.5.1 Privacy Study

With the increased capabilities from the installation and use of pervasive sensor and video networks, serious concerns are raised with regard to privacy and socially acceptable boundaries, creating obstacles to the deployment of such networks [80][81]. However, there currently exists no real data about privacy in relation to ubiquitous systems like the SPINNER media platform and most existing studies are speculative.

Fellow Responsive Environments Group RA Nan-wei Gong has constructed a system that works with the SPINNER system and allows users to control and configure their privacy within the sensor network from both an online web interface (via pre- and post- processing) as well as using a portable privacy badge [28].

This privacy study was run alongside with other applications that use information collected from the distributed network. This allows use-case specific data about people's privacy boundaries to be analyzed. The results of this study can inform future iterations of applications to be more accepted, less invasive, and deployed appropriately.

FIGURE 3-42 Nan-wei Gong's Privacy System and data collected from user tests [28]



### 3.5.2 Awareness Application

A demo application was developed for an event held at the MIT Media Lab with a theme of extending awareness through ubiquitous sensor networks. The event brought hundreds of visitors to the building in which the SPINNER media network resides. This application enhanced the awareness of the visitors

by providing a portal through which they could see through walls and into entirely virtual environments.

FIGURE 3-44 Portal running Awareness Application



The application presented a map of the building on the portal's touchscreen. The map's areas were labelled with information captured from the sensors at that locations, such as the room's activity and if someone was interacting with the portal in that area. Using the sensor data as a guide, the user can navigate the map and view a location. Once a remote location is selected, the portal will display live video and play live audio from the remote portal. The remote portal will display an "on-air" indication notifying the room that video is being streamed to another location in the building.

When no one is interacting directly with a portal, the screen plays canned video content, such as movies appropriate with the theme of the event. Extensions to this application added the ability for wearable devices to select

FIGURE 3-43 On air indication



FIGURE 3-45 Portal displaying a movie



and control the playback of the recorded movie. This allows video content to “follow” a badge-wearer around and playback from where it left off the last time the user approached a portal.

### 3.5.3 Visualization and Mobile Browsing

The SPINNER distributed network generates large amounts of information in the form of sensor data, audio, video, and interaction results. Several applications have been developed that abstract this information into a visual form to convey the key aspects of the data set in an intuitive way.

These applications allow a user a quick view into a complex system or can provide a novel visual experience. The Responsive Environments Group’s Tangible Ubiquitous Media Explorer [85], presented at Ars Electronica by Alex Reben, integrates sound, images, and diverse data sampled throughout the Media Lab into a visual, aural and haptic experience.

**FIGURE 3-46 Mobile Phone SPINNER viewer**



**FIGURE 3-47 Tangible Ubiquitous Media Explorer**



Alex Reben



Several interactive visualization applications have been developed for mobile devices. These applications, such as Manas Mittal's UbiCorder [83], provides in-situ examination of the data from the SPINNER sensor network and helps users make decisions about where they should go or who they should contact based on the current state of the environment.

FIGURE 3-48 UbiCorder



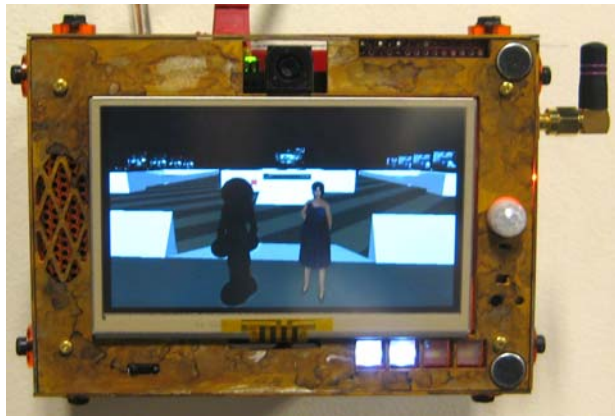
### 3.5.4 Cross Reality

Cross Reality, pioneered by Joshua Lifton [84] using the Plug Point Platform, is the application of distributed sensor networks in which real data is used to enhance the experience of visitors to a virtual world, and vice-versa. This can be used to provide a personalized, ever-changing virtual world based on activities in the real world. It can allow a virtual space to better mimic a real space for remote visits.

An interactive application was developed by Media Lab RA Drew Harry [82] that represented the other end of the portals in the 3D virtual world “Second Life”. Live and cached video could be browsed along with current and historical sensor data at each virtual portal. Users can also stream video

and send messages back from the virtual portal to the real-world one. Applications like this one use a virtual world as a unifying layer to connect different physical spaces.

**FIGURE 3-49** Portal displaying the virtual world view



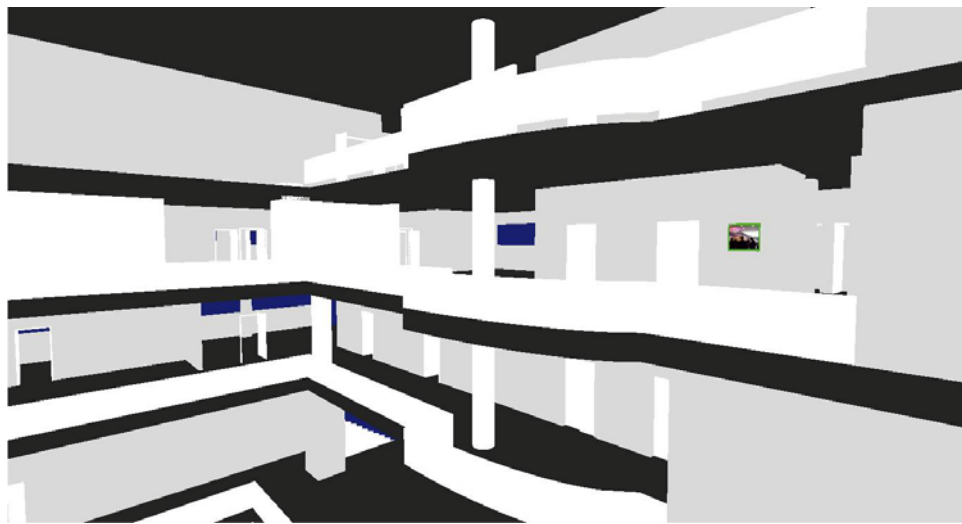
**FIGURE 3-50** Drew Harry's 3D SPINNER-based video library application



*Drew Harry*

Another example of “cross-reality” is bringing the SPINNER portals sensor and image/video data into an interactive 3D building model of the MIT Media Lab via the Blender [86] gaming environment has been explored by fellow Responsive Environments Group RAs Gershon Dublon and Laurel Pardue. Using the SPINNER network to stream data from the real Media Lab to the virtual representation of the Media Lab, provides a stronger connection to what is really happening for visitors to the virtual space.

**FIGURE 3-51** Virtual Media Lab showing a portal into the real Media Lab (center right)



### 3.5.5 Cloud of Images

An application named Cloud of Images was developed and demoed during an event at the Media Lab with the theme of interact - inform - inspire. The Cloud of Images application enables any visitor to the space to take a still image and have it displayed instantly at 40 points in the building. This demo effectively converted the entire building into a digital picture frame displaying up-to-the-minute captured images. In addition to the most recently captured image, RSS news information is displayed. This invited people to move closer to the display, view the RSS feed and the image, and ultimately hit the button to capture their own image and add it to the cloud.

Over the course of the 3-day event, over 1500 images were captured and displayed. As people became more accustomed to the system, it saw some



creative uses such as making announcements to the lab by capturing a written note.

FIGURE 3-52 Cloud of Images showing RSS and the last captured image





# SPINNER VIDEO APPLICATION

.....

**S**tories are the creative conversion of life itself into a more powerful, clearer, more meaningful experience. They are the currency of human contact.

— Robert McKee [95]

The *SPINNER Video Application* chapter describes the design and implementation of the SPINNER Video Application, which utilizes the data from the sensor system to select, collect, and edit video segments into a narrative form.

## 4.1 INTRODUCING SPINNER VIDEO

---

The SPINNER Video Application is the initial application of the platform that motivated its development. The goal of the SPINNER Video Application is to create cohesive videos from actual behavior by uniting wearable sensing with a distributed video capture system.

Using human physical and social actions as an input to a creative process bases the output of the process on reality, ensuring its validity and potential for relevance. Although reality provides this potential, due to the sheer volume of activity (much of which provides no great insight) any actual relevant actions can easily be missed or obscured by overwhelming amounts of non-essential happenings.

The SPINNER Video Application uses the collected sensor data to form and label video segments, which it can then sequence into a cohesive video in a narrative form. The goal of the application is to ensure that the output video is sequenced in a way that considers its acceptability to the intended audience and best represents the core meaning behind the actual events.

The SPINNER Video Application is motivated to have a significant impact on the world of ubiquitous media creation. It allows completely personal video to be created, representing actual events in a storied fashion that can be readily accessed. The creation process is transparent, and real natural actions are the creative input. Through this process, participants can communicate their life to others, and also examine their own life. This brings about the potential for a new community based on story-formed life videoblogs automatically created, and a new avenue for personal self-reflection through examining one's life in an organized narrative form.

Through experimentation with the SPINNER Video Application, we have witnessed the formation of a new space at the juncture of narrative and

reality. We have named this space Story-ality, and it is defined as the perspective of seeing real life events as part of one or more unfolding stories.

## 4.2 NARRATIVE-INFORMED DESIGN

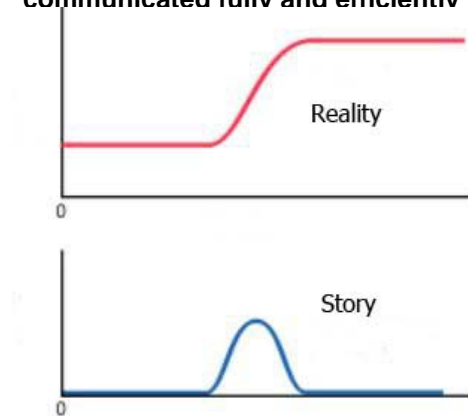
When we talk about cohesion and audience comprehensibility, we are essentially talking about narratives. Narratives are a description of a sequence of events constructed to communicate the ideas behind original occurrences. Narratives chronicle change and can be intuitively reconstructed into the overall meaning of the events that they describe.

An effective, cohesive narrative is one that not only describes the events in a comprehensible and complete way, but also does not include information not pertinent to understanding the original ideas and overall meaning. In other words, it is efficient.

Although there are infinite possibilities for unique stories, it has been well discussed over hundreds of years and it is generally accepted that there are a finite, in fact very few, number of patterns for the overall architecture of an effective narrative. These patterns are concerned with the comprehensibility of the story, in other words, without organizing a sequence of events in an acceptable narrative form, the events will look random, inefficient, and potentially not convey the meaning of the actuality of the occurrences.

Narratologists have thus split narrative into two main components, the overall story structure that allows the story to be read and comprehended by its intended audience, and all the details internal to the story and its characters that make each story a unique experience.

**FIGURE 4-53** Stories can be metaphorically seen as the **first derivative of reality**, which describes only the changes in reality, but in such a way that the overall experience can be communicated fully and efficiently



This dichotomy is expressed succinctly with the screenwriting mantra, “The same, but different”, indicating exactly that the best screenplays utilizes a formula that brings its audience into the story, but with enough uniqueness to make it worthwhile when they get there. When pitching a screenplay, a writer almost always phrases it with a single logline of the form “It’s like XYZ, but with ABC.” An example of this would be, “its like Die Hard, but on a Bus.” This is illustrated by the countless successful movies that fall into the category, “Monster in the House”, such as Jaws and Alien, both of which have exactly the same script beyond the obvious almost find-and-replace differences as far as the monster and the house are concerned. [96]

Vladimir Propp, of the Russian Formalist school of narratology, claimed in 1928 that the formula is everything. He claimed that the story structure is always made of a sequence of up to thirty-one specific entities, and that the details, tone, mood, and character traits are completely irrelevant to the structure and should come from the specific descriptions of each entity. [32] Claude Levi-Strauss has claimed that the structures of stories come from a universal goal to reduce arbitrary data into an ordered, pedantic form illustrating how to progress from awareness of an opposition to its resolution. [97]

Although critics of formalist, structuralist, and reductionist views of narrative [98] claim that the structure comes from the details, they agree that the structure can be viewed separately from the detailed discourse regardless if it is a crafted formula or a by-product of the internal details of the characters, settings, and actions.

The details and the structure can be further discussed as that which is found and that which is created. In many story methodologies, including most documentary styles and the SPINNER Video Application methods, the details are uncovered from reality by the created story structure. In other words, the story is presented in a planned, structured way that engages the audience so

that they can discover and experience the details, moods, tones, emotions, and specific actions of the characters.

The following diagrams, by Kurt Vonnegut [99][100], show examples of the structure of stories. They are drawn in terms of ecstasy vs misery over time. It is important to note, that the ecstasy and misery that is being measured is that of the audience, not of the characters, although in most cases they are related. You can, of course, imagine a story where the audience gets happier as one or more main characters gets more miserable.

**FIGURE 4-54 Vonnegut's Cinderella story structure. Specific story moments are described where a major change in audience experience happens. Note that, although they are specific to Cinderella, they are very easily adapted to describe any event in any story**

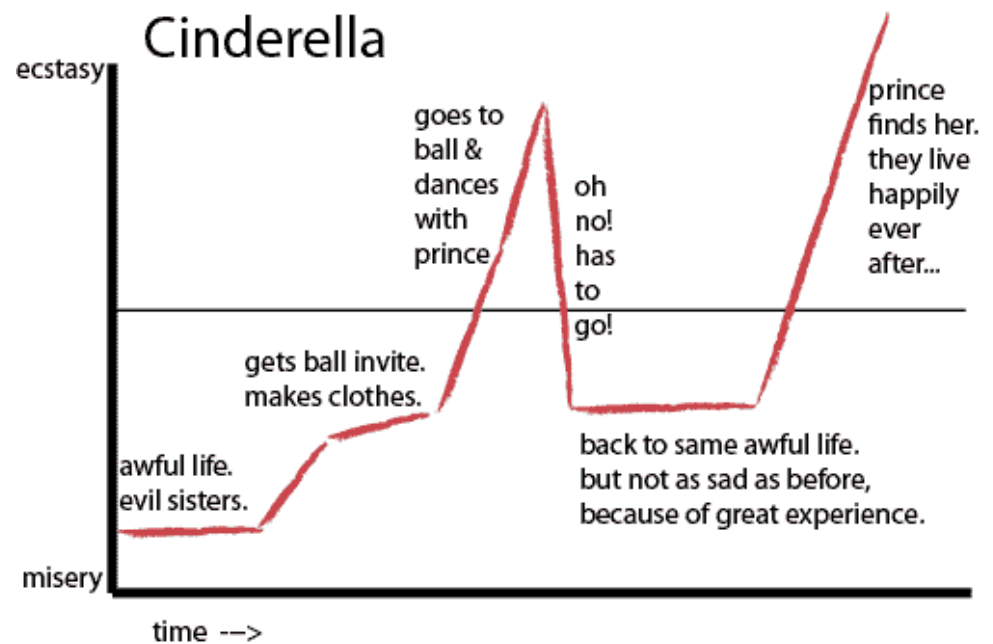


FIGURE 4-55 Vonnegut general disaster story structure



FIGURE 4-56 Standard 3-act story arc

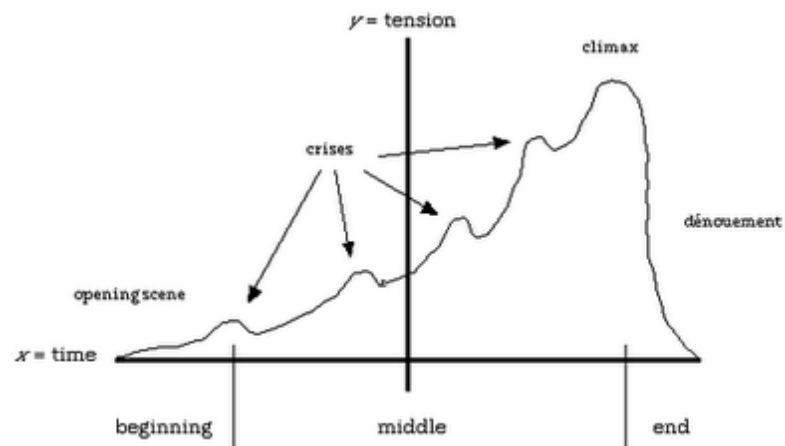
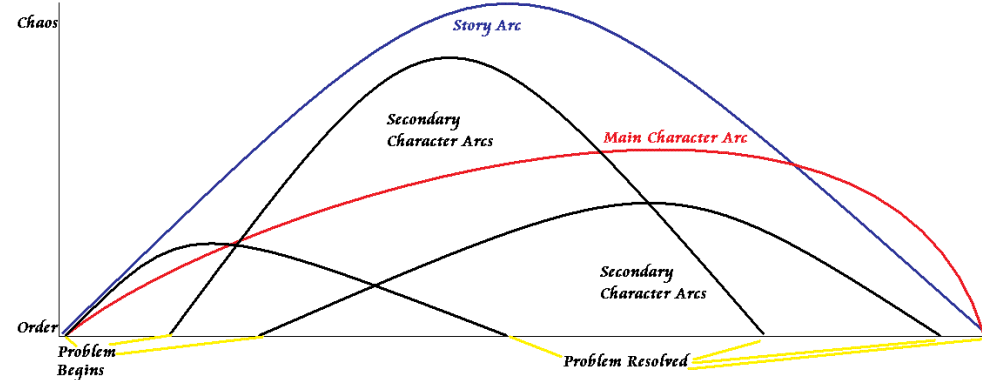




FIGURE 4-57 Example of story arc and character arcs



Since a story can illicit a different reaction from the audience than the current internal thoughts of a character, a story is often split into an audience-specific story arc and one or more character arcs. This echoes the narratological separation of structure and detail.

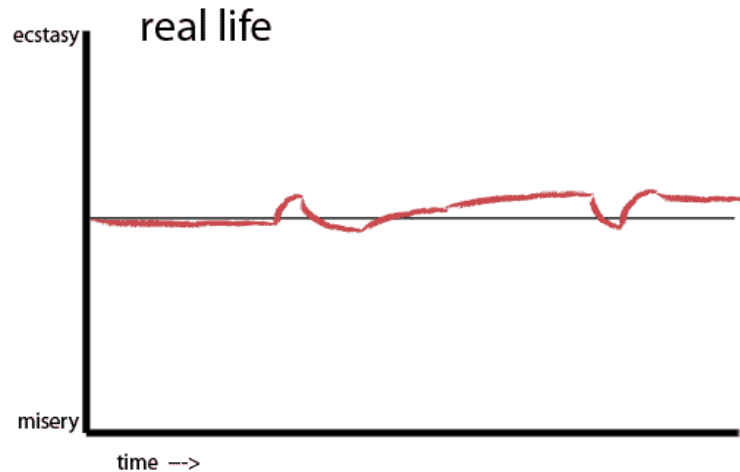
The SPINNER Video application has been designed from the ground up with consideration of how narratives are constructed in order to meet the goal of producing for an audience an understandable, cohesive, and efficient video from real events.

The details of the actions and interactions of the real events are captured with the distributed system of video cameras and microphones. This system captures everything from facial expressions and verbal communications to actual activity and expressed emotions/conflicts. By recording everything, we guarantee that we will have captured rich, detailed content describing a character and a section of time in their life.

However, by capturing everything there will be too much non-essential data for an audience to ever sort through and become engaged with the detailed character content. The goal of the SPINNER application is to not only find the few actions that represent the overall happenings in the character's existence,

but also to ensure they can be presented in a way that the audience can follow, become engaged with, and extract meaning from the overall video.

**FIGURE 4-58** Kurt Vonnegut's Real-Life Story Model, illustrating the sparseness of actual change



The SPINNER video application defines the story structure in terms of story energy. Story energy is a parameter that quantifies the relationship between the audience and the story. At any given time, the audience can feel a positive relationship with the story (happy, hopeful, relieved, etc), a negative relationship (scared, sorry, sad, etc), or indifference. When selecting a clip to include in the final video, the SPINNER Video Application does not need to define the specific emotion felt by the audience, as this will come through in the video. It needs to eliminate clips of indifference, long clips where the story energy does not change, and ensure that the story energy trajectory is followed from clip to clip, otherwise a non-continuous, random-feeling video will be produced, preventing the audience from being able to engage with and understand the story, regardless of the details. On the other hand, if the clips are properly filtered and ordered, even without exact knowledge of their

content, the details will come through. This is owing to the fact that the source material comes from reality and therefore has everything already in it, with reason and detail, and in approximately the correct order. The SPINNER Video application just needs to frame it so that the audience can follow the journey.

The SPINNER Video application achieves this by segmenting the captured video into short clips, labeling these clips with specific details from the sensor system, analyzing the wearable sensor data from the clip to build a local story energy curve of that clip, and sequencing the labeled clips along a story trajectory according to a few input parameters that design the specifics of the target narrative structure.

## 4.3 VIDEO SEGMENTATION

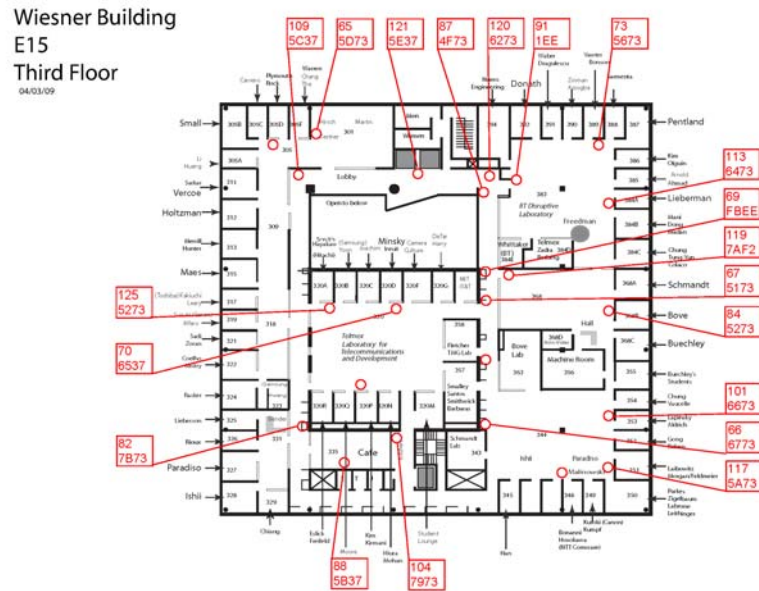
The first task of the SPINNER Video application is to segment the video streams into smaller, manageable clips. Its first task is to break up the video according to whether or not a particular character is portrayed. So each clip should contain a single continuous activity of one character in front of one camera. The SPINNER Video Application takes in a list of characters that it should be looking for during each execution.

The system uses motion detection (both passive infrared and video motion detection) to determine when action is happening in front of a camera. It then uses the location system on the wearable sensors to identify specific participants in front of the camera.

### 4.3.1 Location Engine

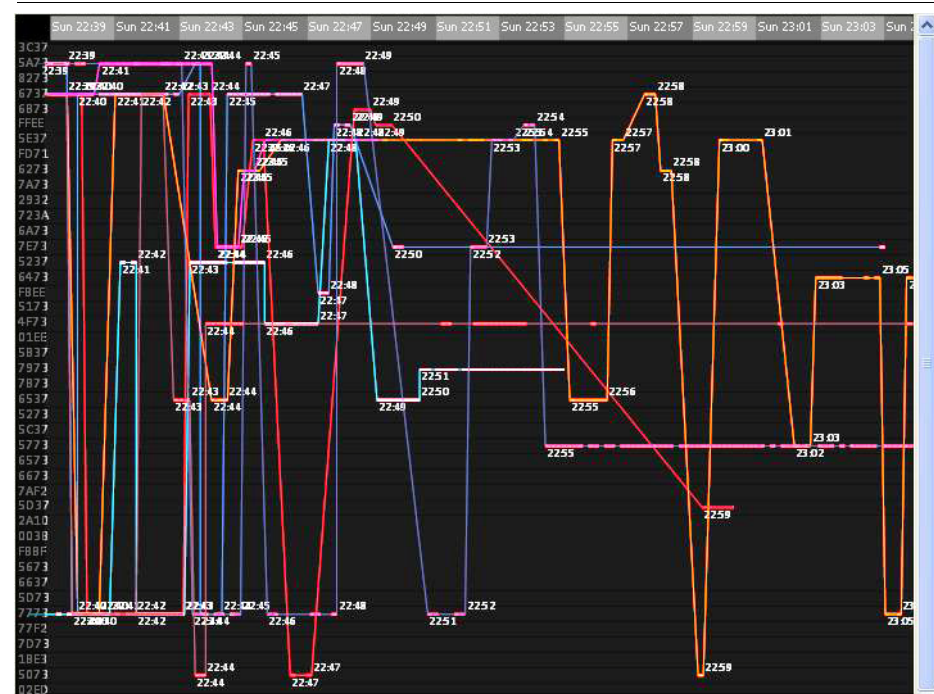
The entire platform has a built in system for tracking the location of people wearing the badge and/or the wrist sensor unit. The location engine uses the always-on portals to receive periodic (every 10s) location announcement pings from the wearable devices. The portals receive these pings along with a link quality indication comprised of analog signal strength and digital bit oversampling error count. This information is then broadcast by the camera on the wired ethernet to other the other cameras, collecting link quality information from all the cameras that received the ping. With calibration, proper camera placement, and the reception of a series of pings from a mobile node in order to remove outliers and errors, the system has been shown to be able to triangulate the location with an accuracy of 3 meters, at least in unstructured informal system observations.

FIGURE 4-59 Map of camera locations during test run

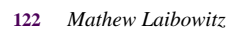


In reality, the layout of the portals is driven by where the action happens and is required to change regularly, making calibration difficult. However, it is easy to determine which camera the participant is closest to with enough accuracy to note if there is a good chance that they will be captured by the camera. In areas where it is possible for the subject to be closely behind a camera, the system uses the signal strength from the rest of the cameras that received the ping (and which cameras did not receive the ping at all) to determine that he/she is behind a camera. For the most part, the cameras are usually placed where a person cannot walk behind them, but where it can happen (particular where the pings can be received through the floor), more cameras are placed in strategic locations to provide the additional information required to differentiate which camera the subject is probably in front of.

FIGURE 4-60 Output of location system during test run



**FIGURE 4-61** Sample wearable node trajectory visualized from collected location data



From this data, a person's trajectory can be drawn. An example of this is shown in **FIGURE 4-61** with the number of concentric circles around each portal indicating the length of time spent in view of the portal's camera.

**TABLE 4-1 Percentage of time spent in front of camera during test**

Participant	Percentage of Time in Front of a Camera
1	70 %
2	62%
3	81%
4	71%
5	68%
6	79%
7	77%

The above table shows the percentage of time spent in front of a camera for each participant during the one hour test. While this can be used to estimate the participant's mobility or "camera-shyness", it is better utilized as a metric for the coverage of the camera system. The average time spent in front of a camera during this test is around 72%, which is acceptable for a research-level distributed system, but missing an average of 28% of the action does leave room for improvement, with the obvious solution of deploying additional nodes.

Our experiment with the SPINNER video application creates videos around a selected main character and uses the location system to ignore video segments that do not contain the main character.

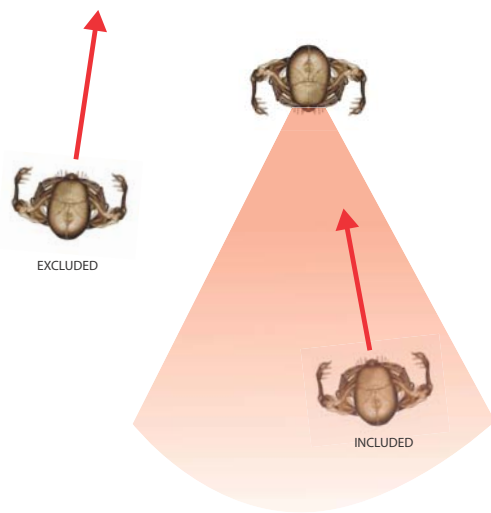
By combining the location information with the camera motion detection, the SPINNER Video application can create a list of video segments per character, with each segment showing that character's continuous time in

front of a particular camera. Should the character leave and come back to the same character, two separate clips are created, removing the time spent away.

### 4.3.2 Social Situation Detection

The clips are further segmented according to whether or not a social interaction is shown. This is important information to the SPINNER video system as narratives are driven, sometimes solely, by changes in social behavior. In addition, the sensor data is analyzed differently during a social interaction than during a non-social activity.

FIGURE 4-62 Diagram of the IR social interaction detection



The wearable devices use an infrared transmitter and receiver with an adjustable viewing angle to detect when two badges are face-to-face and close enough to indicate a social interaction. This system was first developed for the UbER-badge project in 2004 [94] and has undergone extensive development and testing. The effective range of the IR system is adjustable and set to 2



meters for the purpose of conversation detection. The maximum viewing angle is 60 degrees, but this can be reduced by mechanical adjustments in the plastic case. For detecting in a face-to-face conversation is happening, 50 degrees has been shown to be an ideal viewing angle for the IR system. The  $\mu$ Badge sends an IR ping every second which, if continually received by another  $\mu$ Badge, signals an ongoing conversation.

The badge is equipped with the ability to capture speech using a directional microphone pointed directly at the badge-wearer's mouth. Each badge will thusly record a clean audio track of one side of the conversation with minimal interference from ambient sounds or the other person speaking. The badges are accurately synchronized using both the RF Zmat system and the IR social interaction detection system to align the audio with data from each side of a conversation. The badges record the audio with frame synchronization marks broadcast from the cameras. This allows the badge audio to be used in the final product and enables many applications using both distributed camera systems and localized social audio capture.

**TABLE 4-2 Social activity during 1-hour test**

Participant	Percentage of Time Socializing
1	78%
2	83%
3	52%
4	59%
5	61%
6	63%
7	75%

The overall social activity of a person can be calculated by the ratio of the time spent in a conversation to the time not interacting. This parameter is

interesting to compare the participants, as a self-reflective value, or as a test of the conversation detection system.

The SPINNER Video application further divides clips by the social interaction detection. For example, if the camera captures a participant for 90 seconds, of which the middle 30 seconds are spent in a social interaction, the SPINNER video application will split the 90 second segment into three 30 second segments, with the middle segment being classified as a social interaction segment. A study evaluating the detection of social interactions using this system is included in the story energy evaluation in Chapter 5.

### 4.3.3 Labeling of Clips

Once the system has produced a list of video clips according to the subject and the social interaction content, the clips are labeled with the following parameters:

- 1 Location - which camera it was from
- 2 Characters - the main character that the clip was based on, as well as any other characters that happened to be in the clip
- 3 Time - synchronized network time
- 4 Light - the light sensor reading from the camera
- 5 Motion - the motion content from the PIR and video
- 6 Temperature/Humidity - the readings from the camera's environmental sensors
- 7 Social/Not Social - whether or not the clip is considered a social encounter or not
- 8 Audio - the audio captured from all active badges is synchronized with the video from the clip and made available to the SPINNER video application

The wearable sensor data associated with the labeled clips is then processed to provide a local story energy curve as well as specific properties that can be detected by analyzing the wearable sensor data for each clip. This is discussed in Chapter 5. Should there be a long clip with more than one peak or valley of story energy, the clip would be divided around these peaks and valleys.

#### 4.3.4 Evaluation of Clip Detection

The clip detection system was evaluated by comparing it to a human performing the same actions. After a 45-minute run of the system with seven participants, the collected video data was handed off to a non-participant with a list of the seven participants.

The person was instructed to go through the video collected from 30 cameras over the 45 minutes and label the in and out points for each of the seven participants appearing on the video for each camera and the points when a participant went from being by themselves to being involved in a social encounter.

The SPINNER Video Application identified and produced a list of 402 video segments, with labels and synchronized audio, in under 1 minute of time processing the collected sensor data.

The human identified and produced a list of 396 video segments, an error of just over 1%. A quick look at the six clips that the human missed indicated that they were legitimate clips and not SPINNER video application false positives.

It took over THIRTY HOURS for the human to complete this task with virtually the same results as the automated clip detection. The human could not label the clips with the additional sensor information and the timings are less precise. This illustrates that the utility of the SPINNER Video Application

simply as a video annotation tool with enough savings in human resource time to allow the development of large distributed “socially-aware” video camera systems, previously deemed impossible.

## 4.4 INPUT PARAMETERS

---

The clips are then sequenced, further trimmed, and assembled according to a series of parameters that control the SPINNER Video Application. Through these parameters, a user can design the type of final output video that they would like to see assembled from the real events.

### 4.4.1 Story Arc Parameters

The sequencing of video clips, already according to main character, is controlled by the following parameters:

- 1 Story trajectory - the trajectory of story energy versus time of the entire output narrative video
- 2 Curve matching tolerances - the sensitivity of matching the clip's story energy curve with that of the story trajectory. Maximum tolerance will result in a every clip being included, minimum tolerance will require an exact match in both time and value of the clip's curve to the story trajectory
- 3 Social ratio - the ratio of social clips to non-social clips in the final product. Maximum ratio would result in only clips labeled as social being included, minimum ratio would result in the opposite
- 4 Time cheating - the level of allowance of reordering clips to best fit trajectory. Maximum value would allow clips to be placed in any order

that best first the target trajectory, minimum value would require that the clips remain in the order that they occurred

- 5 Flat clip count - clips that have no change in story energy are considered flat. They are matched to flat areas in the trajectory. However, multiple flat clips, each with a different magnitude of story energy, can be sequenced and matched to a section of the trajectory that requires a change in story energy. The flat clip count parameter is the minimum number of flat clips required to match in this way. Should time cheating be set to a high value, the system can pull any of these flat clips to assemble a trajectory to fill in a missing gap. If time cheating is a lower value, then the clips have to be pulled in real order. Setting the flat clip count high will prevent the use of flat clips, except to match to flat areas of the trajectory

#### 4.4.2 Pacing Parameters

The pacing of the output video is controlled by trimming the individual clip length according to the following optional parameters:

- 1 Flat clip duration - For clips with no change in story energy, how long should they be. This parameter can be entered as one value for the whole run, a trajectory that changes over time over flat areas of the story arc, or as specific values for specific times in the output story
- 2 According to labels - The labels, such as light, social, or motion activity can be used to control clip duration
- 3 Change duration - For clips where the story energy changes, the lead and post time of the clip surrounding the moment of the most change (peak or valley) can be entered

### 4.4.3 Music mappings

Music can be mapped to the final output by mapping specific sensor channels to various audio parameters. The simplest form would be to control the volume or tempo of a recorded soundtrack by one or more sensor data channels related to motion activity.

Another possible music mapping is to switch between audio tracks or bring in more instrument tracks when a sensor reading reaches a particular threshold.

In addition to sensor mappings, the audio tracks can be changed by editing decisions made by the SPINNER Video Application. An example of this would be to switch the audio track when the clip changes according to a scripted change in story trajectory or a change between two specific conditions, such as when one clip is social and the next is not.

### 4.4.4 Scripting Specific Points

In addition to mapped parameters, the SPINNER Video Application accepts scripted edit points. In other words, a user could tell the system to make sure to have a transitions between clips at specific times.

The user can also script particular conditions and labels as requirements for certain sections of the final output. For example, a user could say that every clip included in the first 30 seconds has to have a certain light level or be from a particular group of locations. A very common script would be to have non-social clips for the first third, social clips for the second third, and non-social clips for the final third.

#### 4.4.5 Story Energy Mapping

The mapping of sensed and algorithmically-obtained parameters to story energy can be modified according to the target audience, the type of video that is desired, or specific knowledge of the real situation being captured.

The most common mapping used during the tests is mapping motion activity to story energy during non-social events, as changes in the motion intensity are easily observed by the video's audience and indicate a change in the state of the character. While in a social situation, various physical social signals are combined into an energy metric related to interest/attentiveness that is mapped to story energy. Again, these signals directly indicate to the observer what is going on with the characters. The mapping of sensor data into this type of information and ultimately to story energy is detailed in [Chapter 5](#).

### 4.5 FINAL VIDEO OUTPUT

The SPINNER Video Application creates an ordered list of segments. For each segment, the output contains a list of audio and/or video files that contain the information for that segment, along with a list of in and out times for the files to locate the actual clip. Additionally, the segment entry contains a time position for the segment in the overall story.

This information is formatted into a standard Edit Decision List. The EDL file can be sent, along with all the audio and video files, to a non-linear editing software to assemble the final video.



**TABLE 4-3 Example EDL**


---

```

TITLE: bo_knows
FCM: DROP FRAME
001 BL   V   C   00:00:00:00 00:00:38:26 00:00:00:00 00:00:38:26
002 AX   B   C   00:00:34:06 00:00:40:18 00:00:38:26 00:00:45:08
* FROM CLIP NAME: 69_1258947728803_5a73.avi
003 AX   B   C   00:02:48:12 00:02:55:29 00:00:45:08 00:00:52:25
* FROM CLIP NAME: 53_1258947634146_6537.avi
004 AX   B   C   00:00:11:26 00:00:18:14 00:00:52:25 00:00:59:13
* FROM CLIP NAME: 43_1258947773605_7973.avi
005 AX   B   C   00:01:32:10 00:01:43:22 00:00:59:13 00:01:10:27
* FROM CLIP NAME: 49_1258948027783_ffee.avi
006 AX   B   C   00:00:09:14 00:00:20:14 00:01:10:27 00:01:21:27
* FROM CLIP NAME: 17_1258948034328_5d37.avi
007 AX   B   C   00:01:22:29 00:01:43:14 00:01:21:27 00:01:42:12
* FROM CLIP NAME: 39_1258948144364_7773.avi
008 AX   B   C   00:02:04:16 00:02:11:10 00:01:42:12 00:01:49:06
* FROM CLIP NAME: 50_1258948207805_ffee.avi
009 AX   B   C   00:02:28:07 00:02:46:27 00:01:49:06 00:02:07:28
* FROM CLIP NAME: 53_1258948747904_ffee.avi
010 AX   B   C   00:02:19:27 00:02:28:29 00:02:07:28 00:02:17:00
* FROM CLIP NAME: 3_1258948488719_6b73.avi
011 AX   B   C   00:00:00:00 00:00:16:13 00:02:17:00 00:02:33:13
* FROM CLIP NAME: 55_1258949107983_ffee.avi

```

The results of three test runs, each with a different set of input parameters, are discussed in Chapter 6. The output videos are evaluated with a series of questions and by comparison with a randomly sequenced video.

# SENSOR DATA PROCESSING

.....

.....

The moment when the new technologies of photography, film, and the mass distribution of images upset the social and cultural practices of the 20th century is especially striking in Russia, where artistic experiments coincided with great social cataclysms and the search for a new expressivity of the body produced sometimes unparalleled results.

- Oksana Bulgakowa [87]

The *Sensor Data Processing* chapter presents the analysis of the data captured from the wearable sensors. This chapter describes with examples and evaluation the analytical techniques used to interpret the sensor data resulting in information appropriate for the understanding of the captured activity with the ultimate goal of creating a cohesive narrative from the recorded video.

## 5.1 HUMAN SIGNAL PROCESSING

As described previously in [Chapter 4](#), the SPINNER Video Application uses the sensors and devices to divide the enormous amounts of video into smaller, labeled segments. Once the video segments are defined, the sensor data from the wearables over the duration of the segment is analyzed to label the clip with respect to effect the clip will have on the audience, relative to the final video. In other words, the story energy curve is calculated. The actual mapping of raw sensor data and its higher-level algorithmically determined information to the story energy curve is controlled with some input parameters from the user of the SPINNER Video Application. For example, the auteur can decide to map a single sensor, overall motion intensity, speaking emphasis, and/or body jitter to the story energy curve. Every mapping will result in a different video result. The most common mapping to story energy is motion intensity while the subject is alone, and conversational energy<sup>16</sup> as defined below while the subject is involved in a social interaction. This has been arrived at by observing video clips to determine the audience reaction and hand-tweaking the mapping parameters.

This chapter presents the processing of the raw sensor data from the wearable devices into higher-level information. This information, along with the raw sensor data, is made available for mapping to story energy by the SPINNER Video Application, directly labeling the video data for management and browsing, mapping to audio tracks, and for direct mapping to editing decisions.

Pixel processing based vision systems and verbal natural language analysis require a level of domain control over the capture of the audio and video not easily achieved with in-situ camera networks focused on emergent human behaviors. Vision and speech analysis, regardless of if by a human watching the endless amount video content or by a machine, are resource

<sup>16</sup>For this chapter, the term “energy” is used to mean a level of intensity of motion, talking, or listening. In other words, the amplitude of a parameter.

heavy and often cannot run in real-time, making them of little use in a pervasive video capture system.

Perhaps the biggest drawback of video analytics is their limited access to the subject. The cameras in the system provide a remote, third person subjective view, as opposed to an egocentric, body-mounted objective view which, considering the humanity of the subjects, is where the action is. This can lead to significant problems for a vision system, such as occlusion and parallax. Video analytics and speech recognition are often associated with surveillance, as the surveillance industry has driven the technological advancements. Thus these systems are considered invasive, furthered by the proprietary nature of the algorithms and training database, despite their actual lack of subject access making them non-ideal for systems similar to SPINNER.

For the SPINNER system, wearable sensor devices are used to capture on-body low-level sensor data that can be mapped to higher level descriptions of human behavior.

### 5.1.1 Developing the Mapping

The mapping of raw sensor data to more descriptive information has been developed using a top-down approach. The goal of the mapping is to provide information at least as descriptive, if not more so, than can be easily ascertained by watching a video of a real human event. When we watch a video, we quickly pull out obvious information such as who is in the video, where it takes place, if it is a social situation, if there is verbal communication amongst the people in the clip, and if there are any obvious cues as to the activity of the characters.

Beyond these obvious observations, we can often intuitively ascertain the attitude of a person towards their activity and surroundings, the general emotional state of the person, and their relationship with others in an observable social situation. Whether we realize it or not, we observe and

process the body language of other humans. We register things like posture, hand gestures, vocal tone, and facial expressions almost instantly.

Psychologists have studied nonverbal communication, body language, and vocal tone for many decades. [88] The typical methodology for research in this field follows closely to how the SPINNER system is designed. In the typical psychological experiment to study nonverbal communication and/or body language, the researcher will compare the information they obtain from studying the minute details of body motion/expression to the observations of a group of non-biased viewers. These test viewers watch the event either in-person or through video with verbal communication intact. When compared to general observation, the study of gesture and expression yielded significantly more accurate and detailed information about the social relationship, attitude, and feeling of the test subjects; with the obvious drawback of the absence of the actual information contained in the verbal communication. [89] For non-social situations, studying these signals provides otherwise unobtainable information.

The parallelism between the psychological methodology of studying body language and the SPINNER system, combined with the wearable sensors being able to report on body motion, points us to the research in this field of psychology to find the target attributes of the SPINNER system's human signal mapping.

Drawing from psychology to develop a sensor mapping has been validated by Alexander (Sandy) Pentland whose book *Honest Signals* states that “unconscious social signals are not just a back channel or a complement to our conscious language; they form a separate communication network. Biologically based “honest signaling,” evolved from ancient primate signaling mechanisms, offers an unmatched window into our intentions, goals, and values.” [90] Essentially these signals are a physiological response to an

emotional attitude towards a situation and can be seen as a descriptive, universal indication of that attitude.

**FIGURE 5-63 Non-Verbal Communication**



Psychologists have identified many mechanisms for non-verbal communication whether social communication alongside verbal speech or the unconscious “communication” of our internal state. [91] These mechanisms include:

- 1 Gesture - such as hand movements
- 2 Facial Expression
- 3 Eye Contact
- 4 Body position - such as posture, relative orientation between participants in a social exchange, symmetry
- 5 Body motion - mirroring, jitter
- 6 Vocal tone - changes in spectral characteristics of speech, speed of speech
- 7 Conversation dynamics - percentage of time spent talking versus listening
- 8 Space - changes in distance between people, touching, distancing gestures such as crossing arms

While there are many types of non-verbal communication, there is a limit to what can be detected by the SPINNER wearable devices. The devices include a torso-mounted badge and a wrist-mounted watch, as detailed previously in [Chapter 3](#). The devices cannot detect facial expressions, direct eye contact, blinking, touching, and the surrounding space. They have been designed to accurately detect body motion, body position, hand and body gestures, vocal characteristics, speech envelope for conversation dynamics, galvanic skin response, activity level, location, and can be programmed to search for specific actions. During development and testing, much of the

processing is performed offline, but it is a design requirement that they can be performed in real-time on the DSP chip present on all the wearable devices. This allows the devices to store high-level information for peer-to-peer and standalone applications.

It is important to note that the wearable sensors and mappings need not create a detailed window into the soul of the wearer or a complete nonverbal script of an event. All that is required is enough information for the SPINNER video application to be able to select a segment, decide if it should be included in the output video sequence, and appropriately edit and score the segment. The SPINNER video application has been described previously in [Chapter 4](#) with examples of the entire system described in [Chapter 6](#). Additionally, the wearable sensors and mappings should be able to provide enough detail about a video clip for a human to be able to decide if they should view it or not. Although the system does not use attributes such as facial expressions and verbal communication, it captures everything with the video camera and microphones. In general, if there is something interesting in a channel that is not available to the wearable sensors, it will often leave its mark in the channels that are, resulting in the video being included with all the detail intact. The capturing of the video with the sensor data makes the SPINNER system an ideal platform for developing and testing wearable sensor data processing algorithms, as the output of the processing can be compared with observing the video.

### 5.1.2 Summary of Mapped Parameters

The following list is a summary of the implemented and tested mapping targets of the SPINNER wearable sensor data processing system. They are a mix of continuous functions, noncontinuous functions that update their value according to specific occurrences, and tags that just indicate something noteworthy. The SPINNER video application can use the various results of this

**FIGURE 5-64 SPINNER Wearable Devices**



mapping to select and edit a narrative from the labeled video segments. Details of this mapping with data plots from a live scenario are provided in Section 5.2 and the system is evaluated using this data in Section 5.3.

- 1 Talker/Listener - the badge's audio system is designed with the directional microphone facing the wearer's mouth. This limits the amplitude of other voices picked up by the microphone and can identify whether the wearer is talking or listening. This is an important factor for social dynamics but is particularly important as certain gestures mean something different while talking than while listening
- 2 Speaking Energy - indicates how much energy a person is committing to the conversation, how involved they are. It is calculated from the use of emphasis while talking. Emphasis is calculated from the hand motion and the posture. Posture change is detected by dynamics in the pitch of the compass, angular velocity of the gyro, and/or the acceleration due to gravity from the accelerometer. As everyone's posture baseline is different, only changes in posture effect the social energy metric.
- 3 Interest/Listening/Attention Energy- indicates how engaged a listener is with the speaker. It is calculated from hand motion, posture, body position, and body motion while listening. Increased hand motion, particularly if the hand is raised from the side to touch ones face or cross arms, decreases the interest level value. These types of gestures are called distancing gestures. Changes in posture while listening indicate changes in attentiveness, with the more upright the posture, the more attentive the listener. Using the compass heading from both the talker and the listener, the relative body position of the listener to the talker can be calculated. The more square the listener is to the talker, the more interested. Subconscious mirroring of body motion is the final factor in interest level assessment. The various inertial sensors are cross-correlated between the talker and the



listener. When the listener is paying attention to the talker, his/her body motion will tend to mirror that of the talker. Non-mirrored body jitter will indicate the opposite

- 4 Conversational Energy - the combination of speaking energy and listening energy, essentially how much energy a person is putting into the conversation. The calculation can be calibrated to put more or less emphasis on the initial exchange of a conversation depending on how desirable preconceived relationship biases and impressions are to the final value [93]
- 5 Relationship Strength - the conversational energy between two specific participants. In other words, what one person thinks of another, a.k.a. attraction, respect, is approximated by the amount of energy the first person expends in talking specifically to a second person. This measure of relationship strength is read with a high value meaning an active-like, middle value meaning average/indifferent, and a low value indicating an active-dislike. This value can be averaged between two participants to classify their relationship as a unit. Once averaged over multiple relationships this value yields an overall social standing/respect value
- 6 Physical activity/energy - the physical activity parameter(s) is/are calculated from the activity envelopes of 10 channels of inertial measurement, 5 on the hand, 5 on the body. Depending on the application, these can be reduced down to as few as one parameter indicating the subjects's overall activity level. An activity classifier application utilizing Hidden Markov Models or Neural Nets, would likely keep all 10 channels intact. The SPINNER application reduces it to 4 parameters, translational activity and rotational velocity for the hand and for the body. This reduction is calculated by the vector addition of the 3 axes on the accelerometers and the vector addition of the 2 axes on the gyros. The activity envelope is then calculated for

each of these 4 parameters by a windowed variance. These envelopes could be combined should further reduction be necessary

- 7 Internal activity/energy - the galvanic skin response is mapped directly to an internal activity parameter if there is little to no physical activity. While it is always difficult to determine what a change in GSR indicates, if there is physical activity, it is completely meaningless. During a non-social period of low physical activity, the GSR can indicate something of interest. During a conversation, the GSR does not report anything not already seen by the gestures and audio
- 8 Audio analysis - the wearable badge can capture full spectrum audio and has an onboard DSP chip making real-time audio analysis possible. Offline audio processing provides endless possibilities for analysis. Sharp frequency shifts indicate excitement and can be added to the social energy metric. Also, vocal tremors indicating stress can be detected. Unfortunately, most nonverbal speech classifiers are not universal and require subject-specific training

## 5.2 MAPPING DETAILS AND TEST RESULTS

This section illustrates the mapping of the raw data from the wearable sensors into the above listed parameters. The data presented is from a live scenario, where seven participants each wore a badge and a wrist sensor for one hour while within the coverage area of the SPINNER distributed camera system. The test run presents a slightly compressed view of normal behavior as the subjects admitted to engaging in more activity and social interactions during the test than they would during a normal hour of daily life. This ultimately provided a more aggressive test through additional stress on the system. This also had the added effect of the participants quickly becoming comfortable with the devices and behaving more naturally during activity and interaction, even if the gross quantity of these events was heightened by the test scenario. This was easily confirmed by reviewing the captured video and by performing exit interviews.

The SPINNER wearable devices were designed with enough features to support general applications in on-body/social sensing, as well as designed with specific features driven by the target application of understanding video in order to create narrative output. Many of the sought after parameters are mapped directly to data channels on the devices, or arrived at with basic calculations. The following sections present the data and the mappings in growing order of analytical complexity.

### 5.2.1 Social Interaction Dynamics

The signals unconsciously emitted during a conversation are more easily interpreted than non-social activity. [92] Once a social-situation is detected, it also becomes important to detect when the person is talking and when the person is listening, as the analysis of the social signals is different for these two states.

The badge is equipped with the ability to capture speech using a directional microphone pointed directly at the badge-wearer's mouth. Each badge will thusly record a clean audio track of one side of the conversation with minimal interference from ambient sounds or the other person speaking. The badges are accurately synchronized using both the RF Zmat system and the IR social interaction detection system to align the data from each side of the conversation.

The audio envelope from each badge is calculated in real-time. Since the badges have a direct line to a single person's speech, it is easy to use the envelope to determine if that person is talking or listening. Each badge also has a digital noise floor removal algorithm coded using the onboard DSP. This is comprised of a notch filter around the peak low frequency. The noise floor has proven to be pretty consistent from badge to badge allowing the use of tested coefficients and requiring minimal processing power.

The plots on the following pages were taken from two badges after the IR system indicated a conversation. The duration of the conversation, bounded by the IR and the first and last audio peaks during the period of IR detection, is shown with the large red box. To test the noise cancelation, one out of the seven nodes had the noise cancelation disabled. This conversation, between participant 6 and participant 7 (noise canceling disabled) was selected to compare the results with and without the noise canceling algorithm. The first plot shows the waveforms as they were recorded on the node, decompressed for print.

The second plot shows the energy envelope. The two curves on each plot show the sampled energy level and the energy level calculated from the audio waveform. The sampled energy level is a running windowed average of the absolute value of the sampled amplitude. The calculated one is a windowed bandpass filter. A threshold on the filtered energy is mainly used to determine talking and listening, and a second threshold the more accurately timed sampled energy parameter precisely adjusts the moment that talking stops.

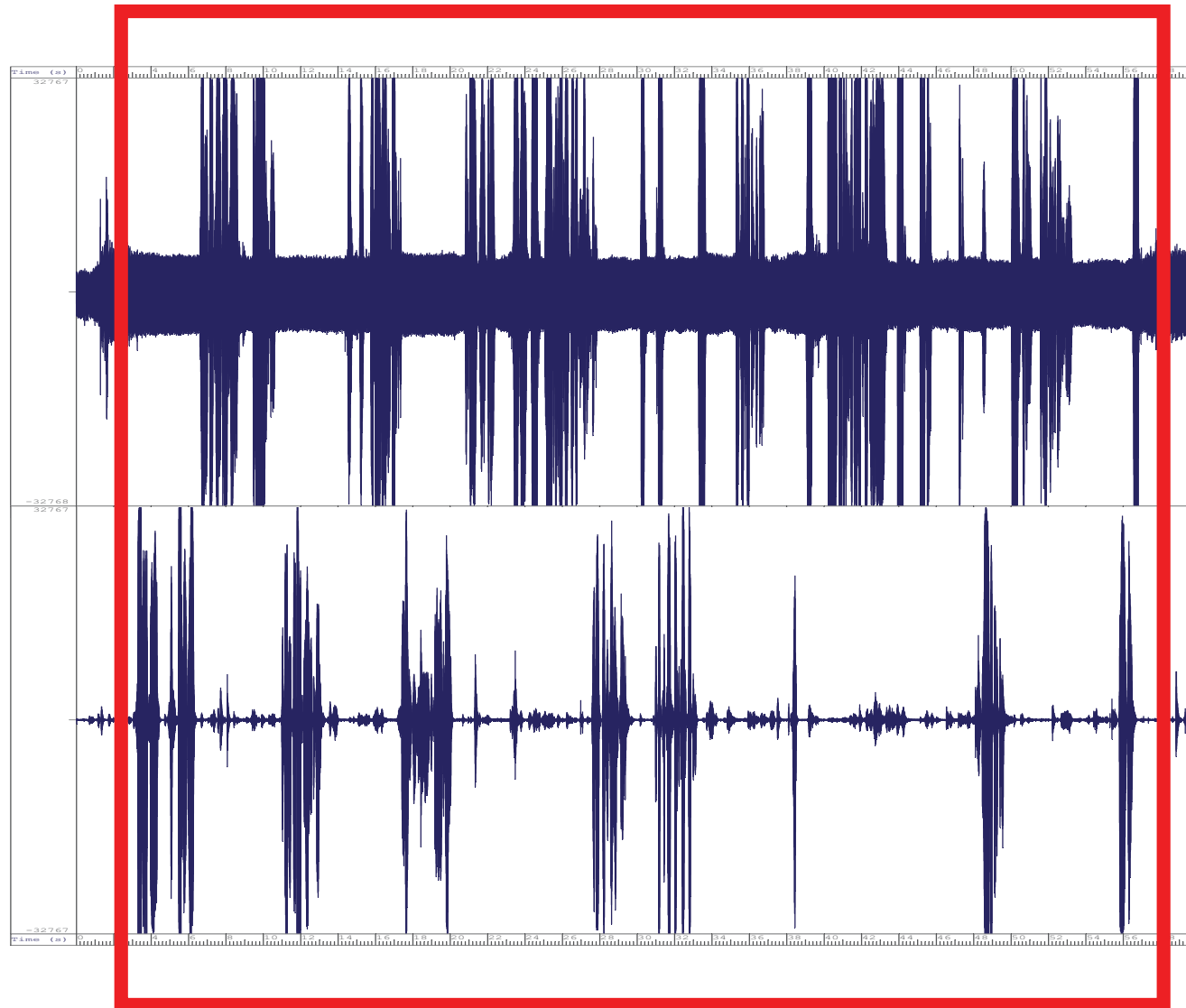
In addition to determining when someone is talking and when they are listening, the badges record the audio with frame synchronization marks broadcast from the cameras. This allows the badge audio to be used in the final product and enables many applications using both distributed camera systems and localized social audio capture.

The plots below clearly show that the noise canceling removes the noise floor from the waveform. This translates to a cleaner signal for the sampled energy envelope. There is a delay associated with the noise cancelation that is not shown as it works on a window of samples. For the filtered energy envelope, the noise is less of a problem as the filter reduces audio at the noise frequencies, inherent to its calculation.

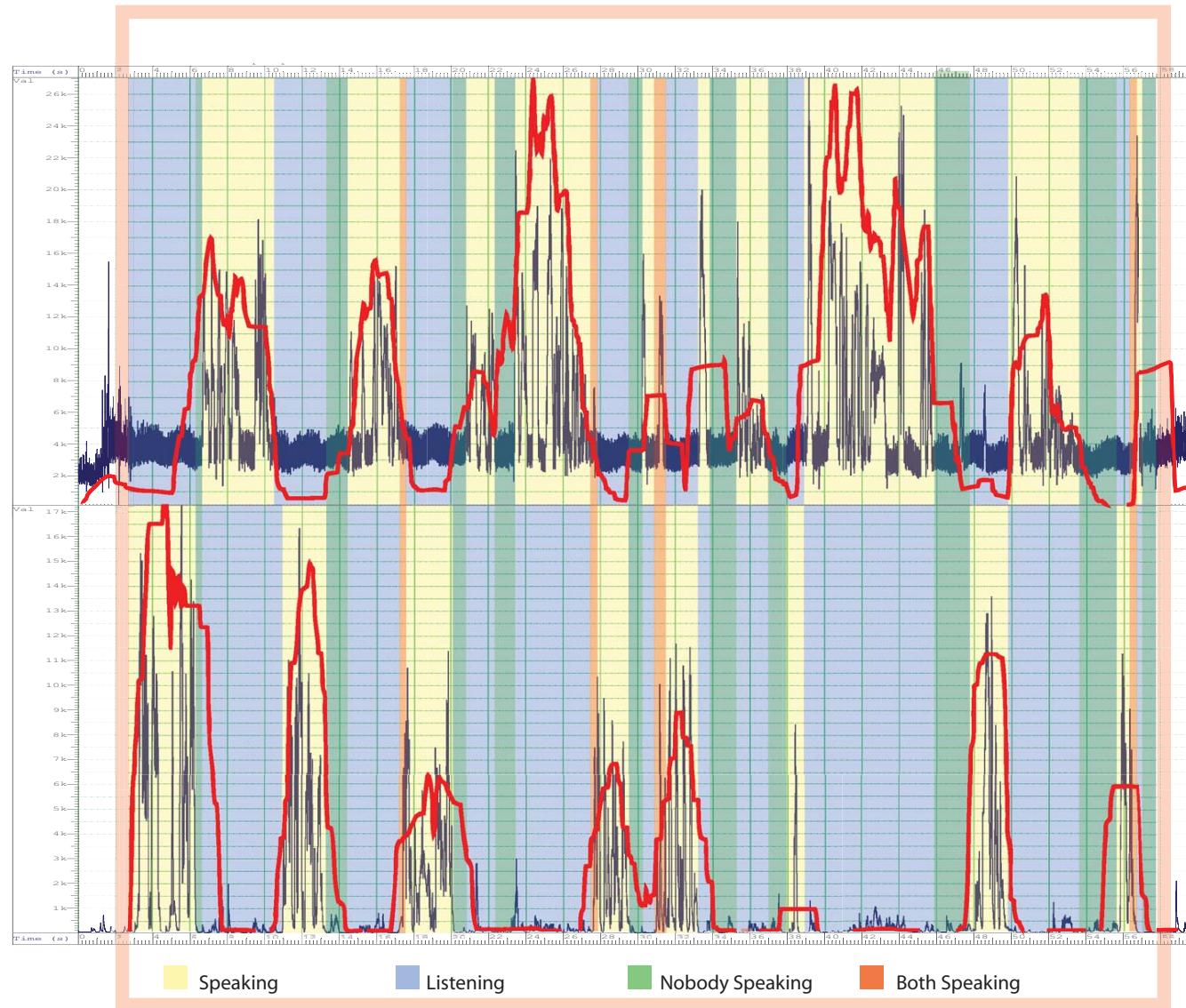
Since these are battery-powered devices, the noise is not 60Hz hum. Looking at the audio spectrum, this noise is most likely hiss from the HVAC systems and moving air being amplified by the high gain of the microphone input amplifier. Since the noise shape is fairly constant on all badges, the noise can be sampled and analyzed to best create the appropriate filters for the noise cancelation system.

The audio envelopes are also used for the social signal analysis discussed in Section 5.2.3 and the audio waveforms can be used for tonal analysis, as discussed in Section 5.2.4. The speaking/listening detection accuracy is evaluated in Section 5.3. It is point of future work to utilize additional details that the badge can easily provide for the social signal analysis, such as speech pauses, interruptions, and speed.

**FIGURE 5-65** Audio waveforms from the badges of participant 6(top) and participant 7(bottom), displaying a conversation from the 1 hour test



**FIGURE 5-66** Audio energy envelopes illustrating the conversational dynamics between participants 6 (top) and 7 (bottom)





### 5.2.2 Activity

The intensity of activity is arguably the most important parameter for the SPINNER system. The level of activity is easily observed when watching video and periods of little activity or no change in activity intensity rarely, if ever, convey visually anything of significance. The goal of the system is to visually tell a story about events in the life of our subjects. Stories are constructed by sequencing events of action and change which can be used to reconstruct the meaning of the original events without the thickness of everything in between the moments of change and action. In essence, a story is the first derivative of reality.

The activity parameters indicate an action resulting in some kind of a change, be it a change in current task execution, mental state or attitude, and/or social situation. These changes present themselves in an observable and physical fashion. If they have no physiological representation, then they will not be detected by the wearable sensors, however, in this case they will not be useful as they also will not be visible to viewers of a video clip.

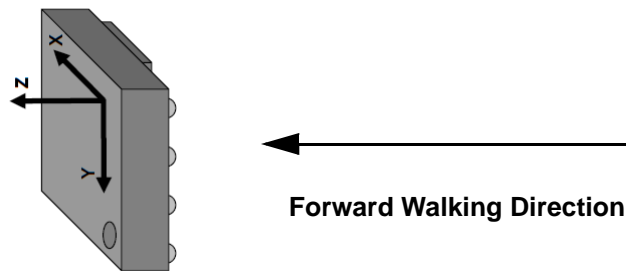
The SPINNER wearable sensor devices have been designed to accurately detect motion with the inclusion of inertial sensors mounted on the wrist and on the torso. Each of these devices has a 3-axis accelerometer and a 2-axis gyroscope. The badge device also has a 3-axis compass reporting absolute angular orientation. The compass is used for various social signals but is generally too slow and not as accurate as the inertial sensors in detecting changes in motion to be used in the activity calculations.

The most useful representation of activity for the SPINNER application is a scalar quantity indicating the level of change in physical activity. This can be calculated with a sliding window variance, centered around the current sample to minimize delay. Variance is a measure of the variation of the signal



by averaging the deviation from the mean for each sample in the window. This is directly correlated to the level of change in activity. The windowing also low-pass filters the data, which is necessary when converting data at 100 samples per second into changes at the human scale. The size of the window is an important argument as it determines the sensitivity of the algorithm. For example, if a person turns their hand 90 degrees at a steady pace, it will present as a baseline shift in one axis of the accelerometer due to a change in the inclination related to gravity. A smaller window, despite being more accurate during periods of high activity, could miss this change completely as each window only has a single sloped line. A larger window can contain the entire transition resulting in a short activity spike. It is often not desirable to have a slow change in inclination considered as activity, but the window size is considered as an input variable to the system when designing the test run. Actually, for a torso mounted sensor, changes in inclination, even fairly slow, can indicate a major change in state.

FIGURE 5-68 Badge Accelerometer Axes



The following two plots show a ten minute stretch of actual sample data taken from Participant 3 during the one hour test run. As ten minutes of data contains 60000 sample points, these plots are fairly zoomed out and provide an overall script of the ten minutes.

FIGURE 5-67 Badge Gyro Axes

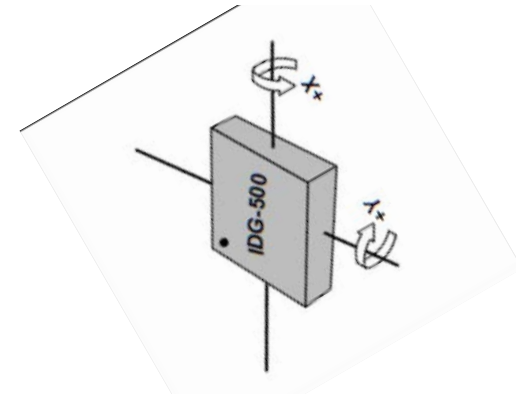


FIGURE 5-69 Example activity plot from the Body Accelerometer Magnitude

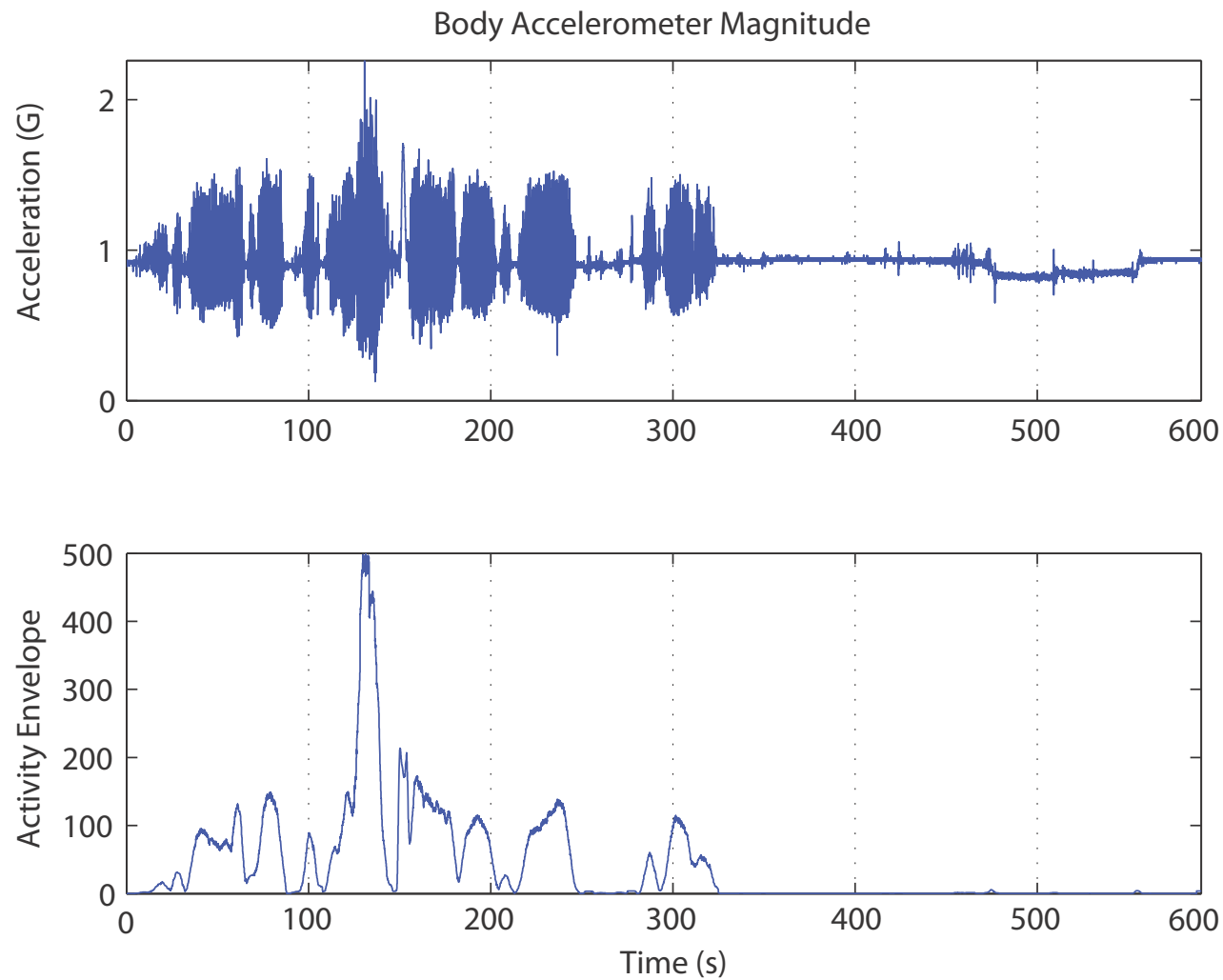
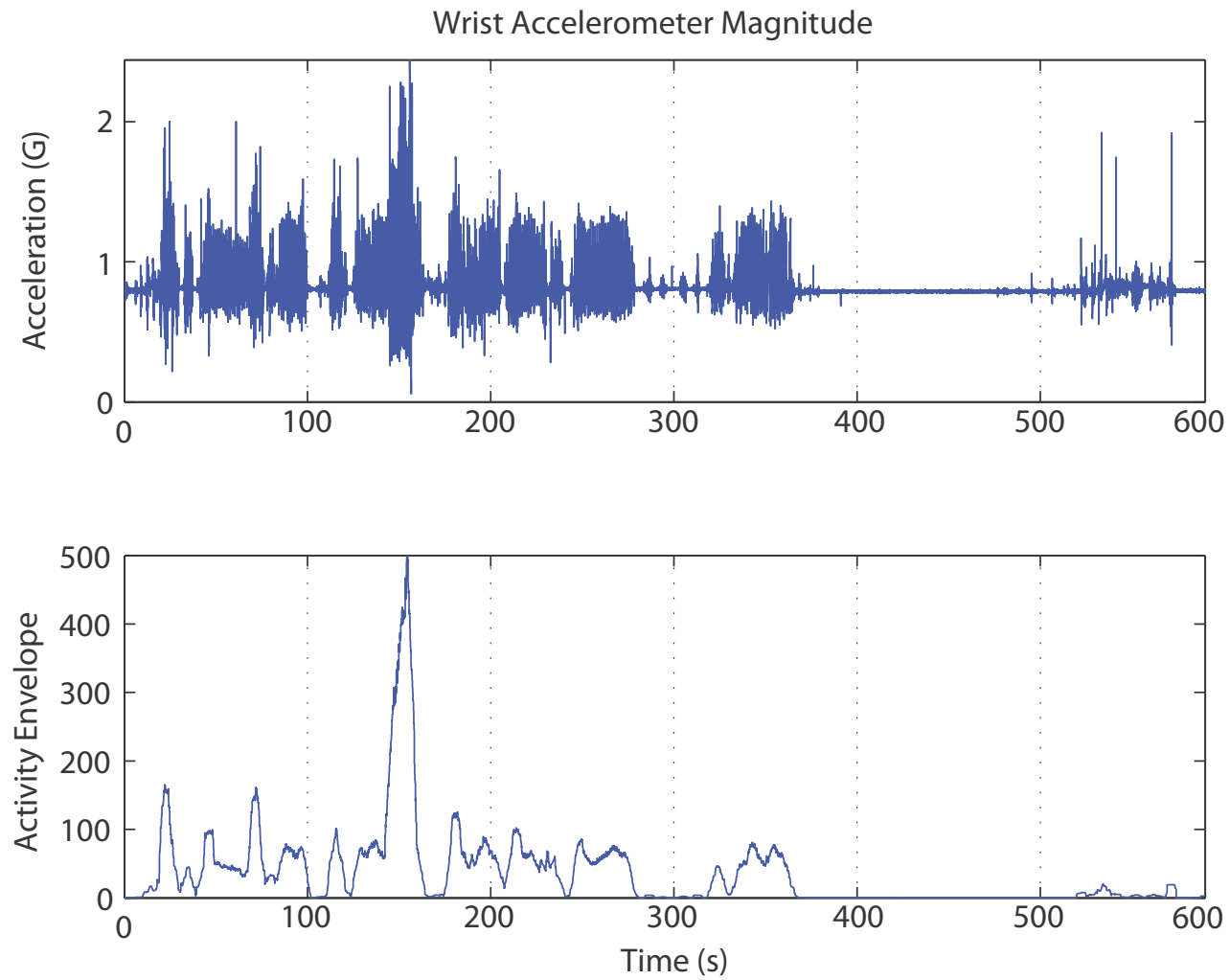


FIGURE 5-70 Example activity plot from the Wrist Accelerometer Magnitude



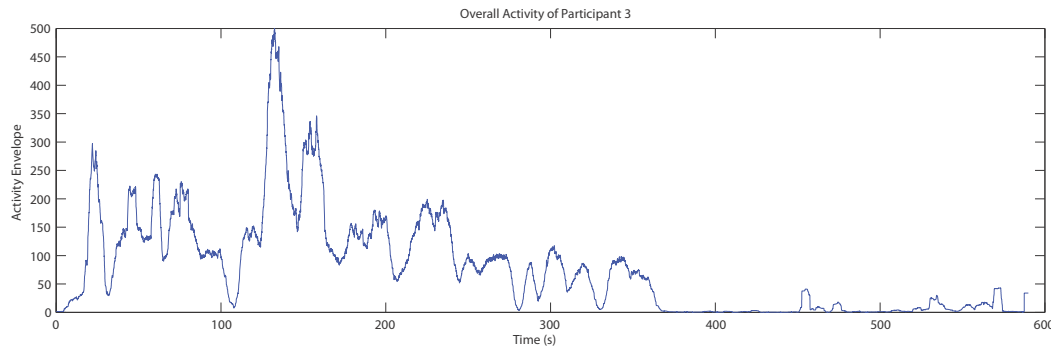
Each plot is paired with its calculated activity envelope. Activity magnitude is used directly by the SPINNER system to determine points of interest during times of no social interaction. It is also used for many of its higher level mappings, including the social signaling discussed in Section 5.2.3.

During this ten minute stretch, Participant 3 spent the first five minutes walking around, followed by five minutes of sitting. At approximately eight minutes in, Participant 3's body angle changed to include more forward lean. At the same time, the wrist activity increased. This combination of events indicates that Participant 3 began to use her computer. A quick review of the synchronized video verifies these claims.

Having ten individual parameter curves provides a level of detail not usually required for the SPINNER video application. At most four activity parameters are needed, body translation, body rotation, wrist translation, and wrist rotation. The last four plots show this reduction using vector addition. When designing a SPINNER application test run, four thresholds are needed to determine the sensitivity to activity, which ultimately determines the length of the final video output.

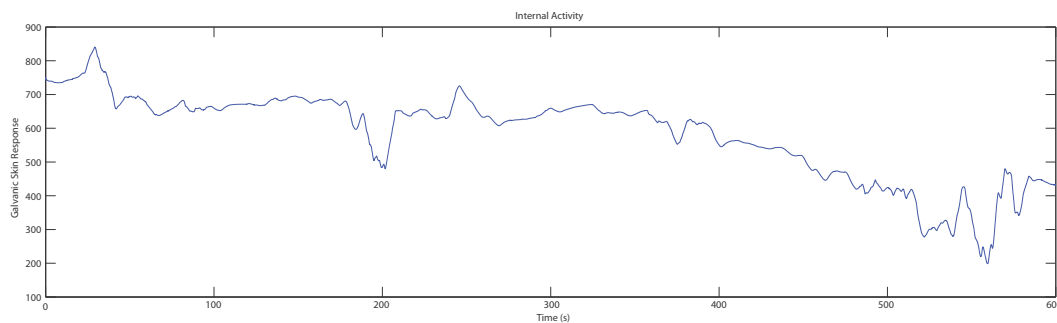
Since the activity envelopes are normalized and all ranged properly, they can be averaged to further reduce the number of activity curves. As the social signal mappings differentiate between the motion of the hand and the motion of the body, the SPINNER applications were designed with the above four activity curves. The curve shown in **FIGURE 5-71** is an example of a single-value activity metric.

**FIGURE 5-71 Overall Activity Envelope**



The wrist sensor device is also equipped with a galvanic skin response sensor. The sensor works by checking the skin conductivity to assess the moisture content. The moisture content in the skin fluctuates according to external temperature, physical activity, and all things being equal, emotional arousal.

**FIGURE 5-72 GSR Response of Participant 3**



The curve shown in **FIGURE 5-72** plots the GSR response from Participant 3 during the same time period as the above physical activity curve. The GSR, as expected, tracks with the physical activity. In the case where the GSR rises above a threshold in a period of little physical activity, the SPINNER system will treat the internal activity as if it were physical activity.

The GSR value is recorded but ignored during periods of high physical activity. Similarly, it is recorded but ignored during social interactions, although the usefulness of GSR in social situations is constantly under test [103].

### 5.2.3 Social Signals

During a social interaction, we express significant amounts of information as to our situation, internal states, and personality. Stories, particularly stories created visually, often focus around the relationships between the characters and the relationship between the main character and the social world as a whole. The life of the character is accurately represented through a chronicling of changes in relationships.

The SPINNER application creates a parameter ranging from a very negative relationship (i.e. active dislike) to a very positive relationship (i.e. active like). With the center point being no relationship (they have not met) or indifference. An overall social standing metric can be obtained by averaging the individual relationship values from all the social interactions of a particular character. This relationship strength parameter is sometimes referred to as respect, social level, or level of attraction.

The relationship strength parameter is calculated from the level of interest/attention one person pays to another and the energy used when communicating. The interest level determines if the relationship value will be positive or negative. The interest level is then combined with the amount of energy expended in the conversation to determine the strength of the relationship, be it negative or positive. The parameter is calculated for each side of the relationship individually. Averaging the scores of everyone's relationship with a single person will yield the social standing value of that person. Averaging the score of a single person's relationship with everyone

else will yield a general social involvement statistic. The relationship scores are continually calculated, as it is important to not only know the levels, but also to know precisely when a relationship has changed and in which direction.

Psychologists have shown that in social encounters the body language during the initial dialog exchange contains information strongly related to either preconceived biases should the person be already known or physical first impressions. The calculation of relationship strength can be weighted more or less during the initial interaction, should these impression factors be desirable to the final product.

#### **5.2.3.1 Speaker and Listener Intensity**

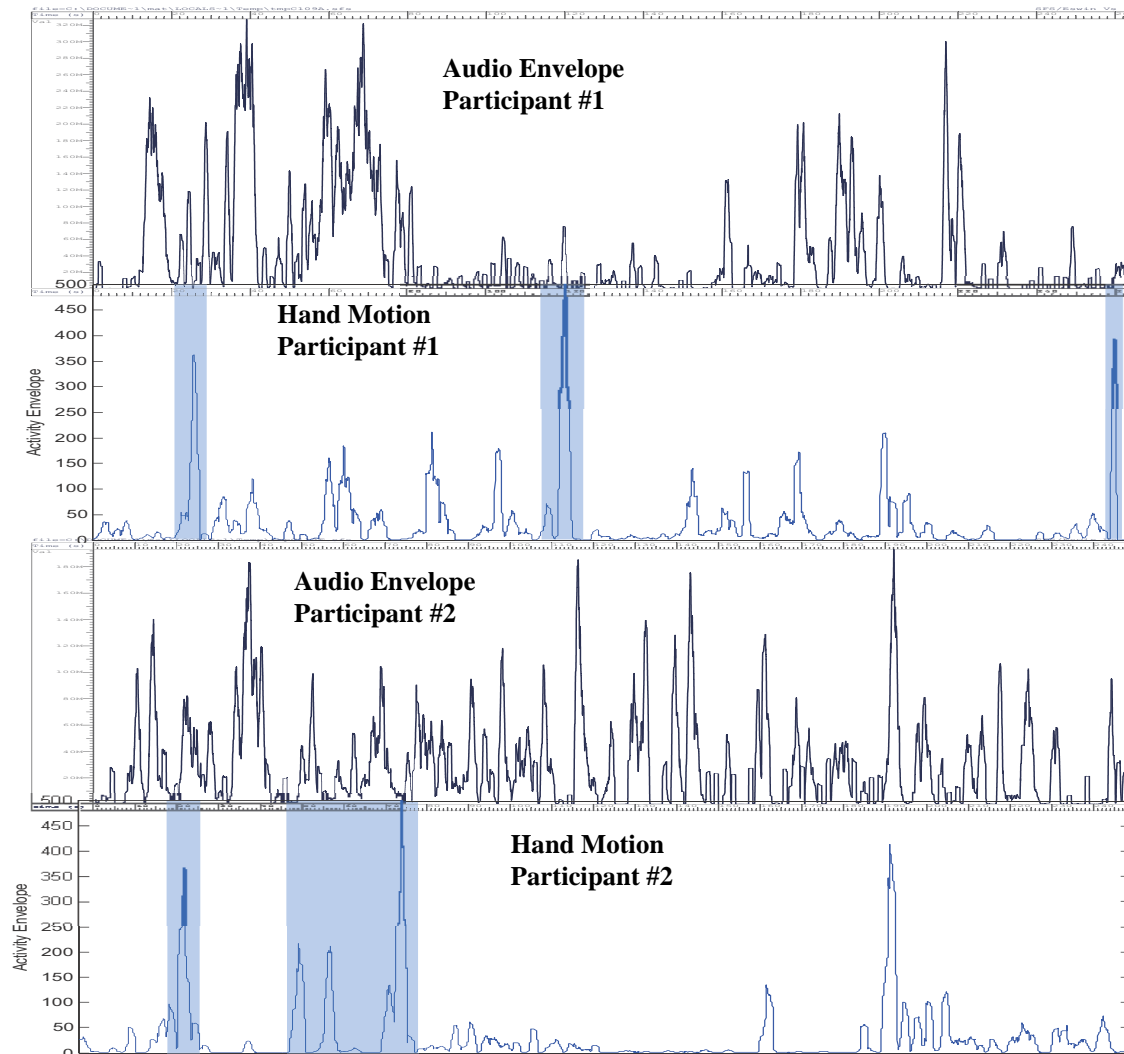
In order to calculate the relationship scores as defined above, we need to first calculate the intensity of speaking and listening. These parameters can also be used directly by various applications. While in a social situation, body language becomes well defined, active, and universal to the point of being able to determine the interest level shown by a listener and the emphasis shown by a speaker.

The SPINNER system calculates the listener level from a weighted average of hand motion, arm angle, posture, body orientation relative to the speaker, mirroring of body motion, and non-mirrored body jitter.

An increase in hand activity while listening, detected directly from the sensors and converted to an energy envelope as described in Section 5.2.2, indicates a decrease in attention towards an active dislike for the speaker.

When the arm is raised off the side, this effect is magnified as this type of gesture subconsciously puts a barrier between the speaker and the listener. The arm angle is provided by the windowed mean of the wrist accelerometer value, i.e. the acceleration due to gravity.

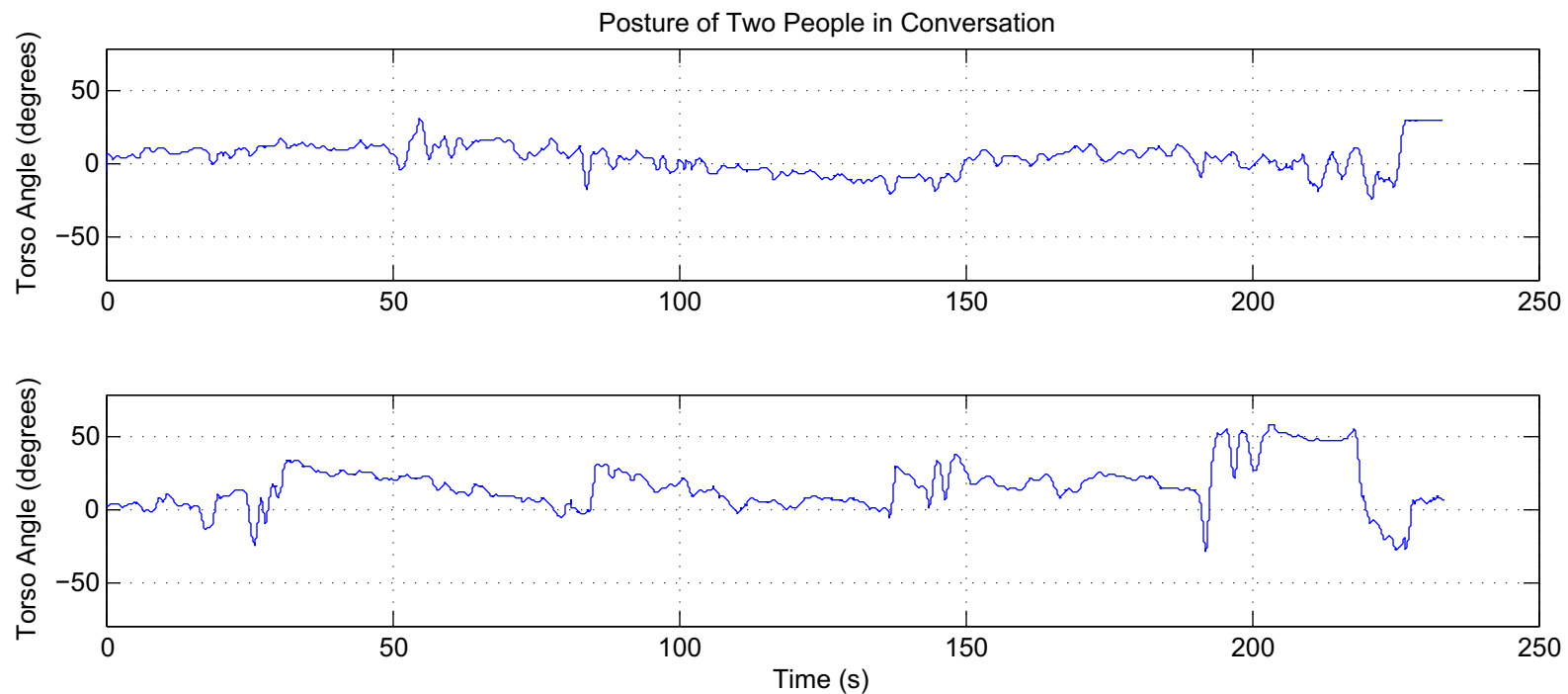
**FIGURE 5-73** Hand activity during listening highlighted for two participants in a conversation, audio envelope and hand motion shown, data taken from actual natural test situation. Shaded areas are time periods automatically selected when hand motion is high and speech envelope is low. These areas are considered as times of low attentiveness due to hand motion while listening





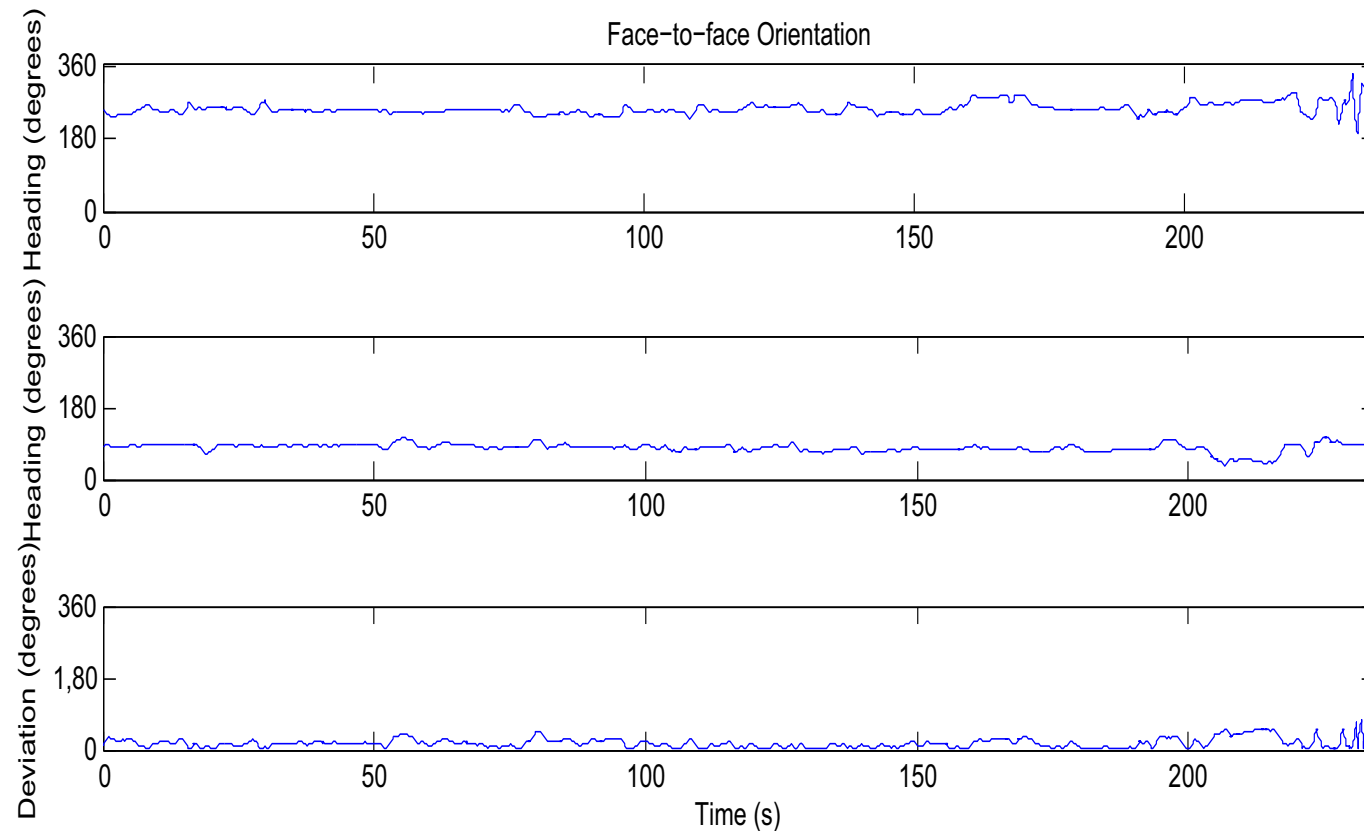
Posture is directly related to attentiveness, even at a very small scale. While large baseline posture changes indicate a major change in listener intensity, even small jerks up and down signal changes in attention. The badge compass' pitch axis directly measures the posture relative to the local Earth's magnetic field vector. For more detail, the compass can report the general orientation and the mean accelerometer and the dynamically integrated gyro can all be used together to accurately measure the posture angle. The posture angle can then be derived into the posture velocity, which can indicate posture slumps and posture inflations.

**FIGURE 5-74** Posture angle of two conversation participants, one shows a smooth bobbing with a generally steady trend, the other shows significant bobbing with an emphatic finish to the conversation



Using the compass heading, the absolute orientation of the torso can be ascertained. By comparing this to the compass heading reading from the person speaking, the relative orientation can be calculated. The more square the two people are, the more the listener is seen to be attentive.

**FIGURE 5-75** Compass angles of speaker and listener, and their relative angle, these two participants show a very low deviation angle indicating attentiveness. Towards the end of the conversation, the deviation angle shows motion, indicating a change in conversational state, most likely from some sort of goodbye gesture



The final factor used to determining interest and attraction level is the level of body motion mirroring presented by the listener. As a person pays more attention to another person, they subconsciously tend to align and synchronize their body motion. Whereas if they are not paying attention they will tend to jitter in no way similar to the speaker. The level of mirroring is calculated by taking the windowed cross-correlation between the speaker and the listener for each of the five inertial sensing channels in the badge. These five values are then averaged to get a general mirror score. Should the mirror score be low, then the activity level of the body is considered jitter and is therefore a negative contributor to the attention level parameter.

The cross-correlation data is viewed in the form of a three dimensional plot. The cross-correlation value ranges from -0.8 to 0.8, A high positive score indicates a high-level of similarity between the two signals, and a high negative score indicates a high-level of opposite similarity, in other words a mirror image. Both straight similarity and opposite similarity count as sympathetic body language. In the cross-correlation plots, dark red and dark blue indicate high correlation, and the lag value can tell you who is the leader. The values for each time window (column) are averaged into a mirroring score.

**FIGURE 5-76** Cross-correlation between two people in a conversation, known to be engaged in the conversation and mirroring to some extent

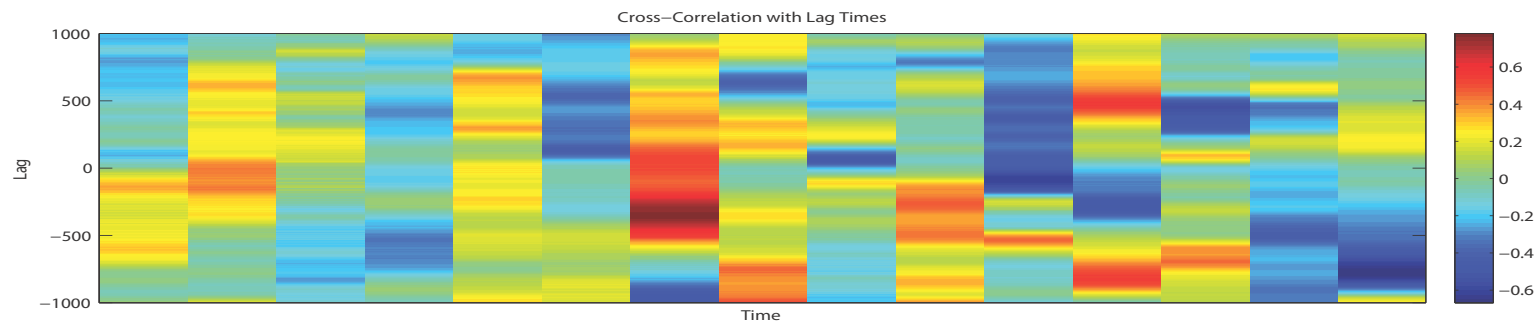


FIGURE 5-77 Mirroring score of above cross-correlation calculated by averaging the columns

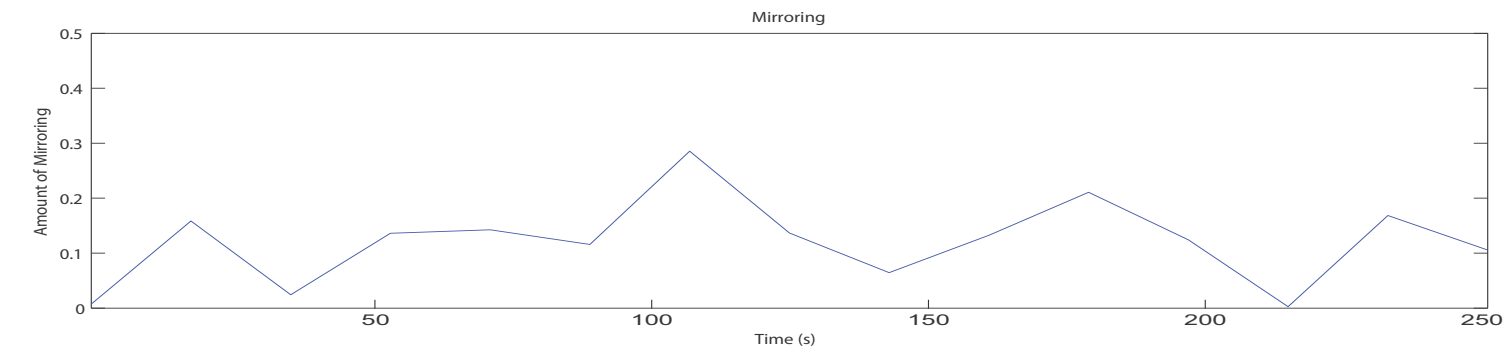
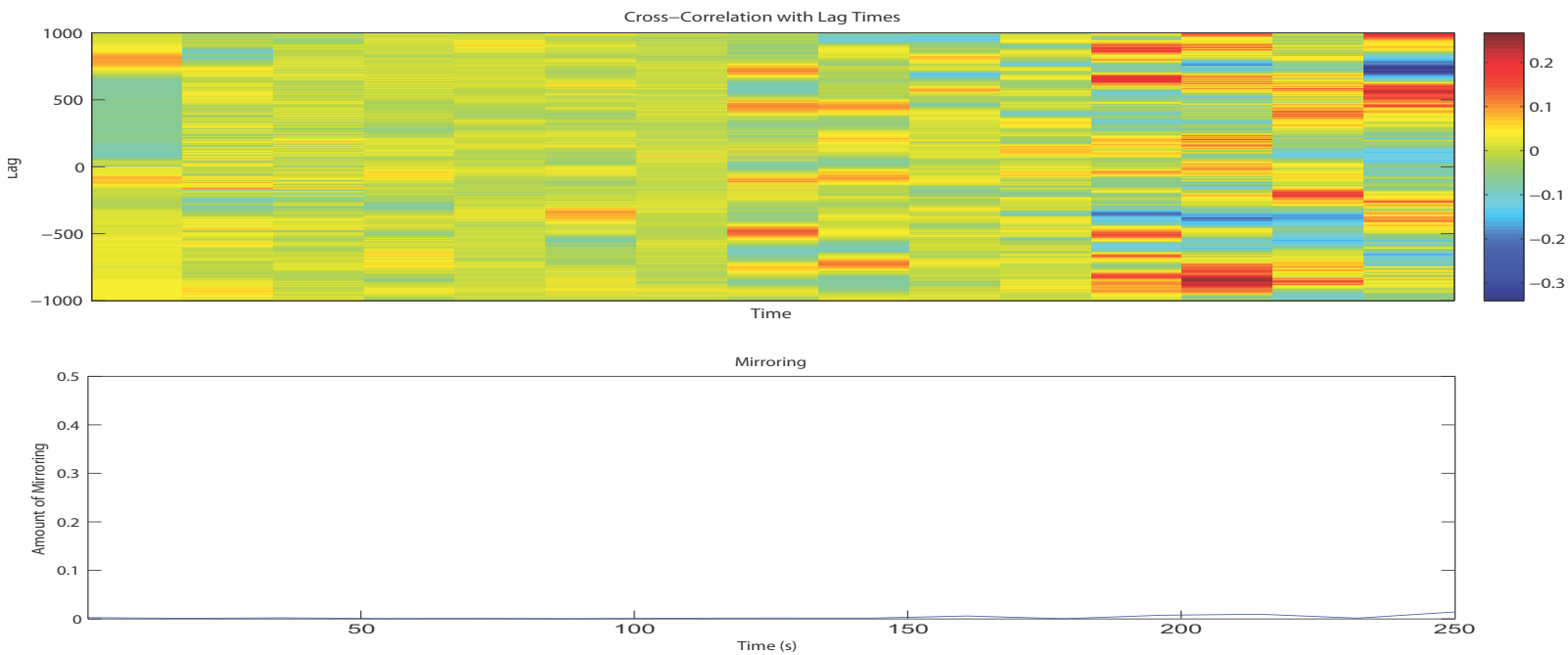
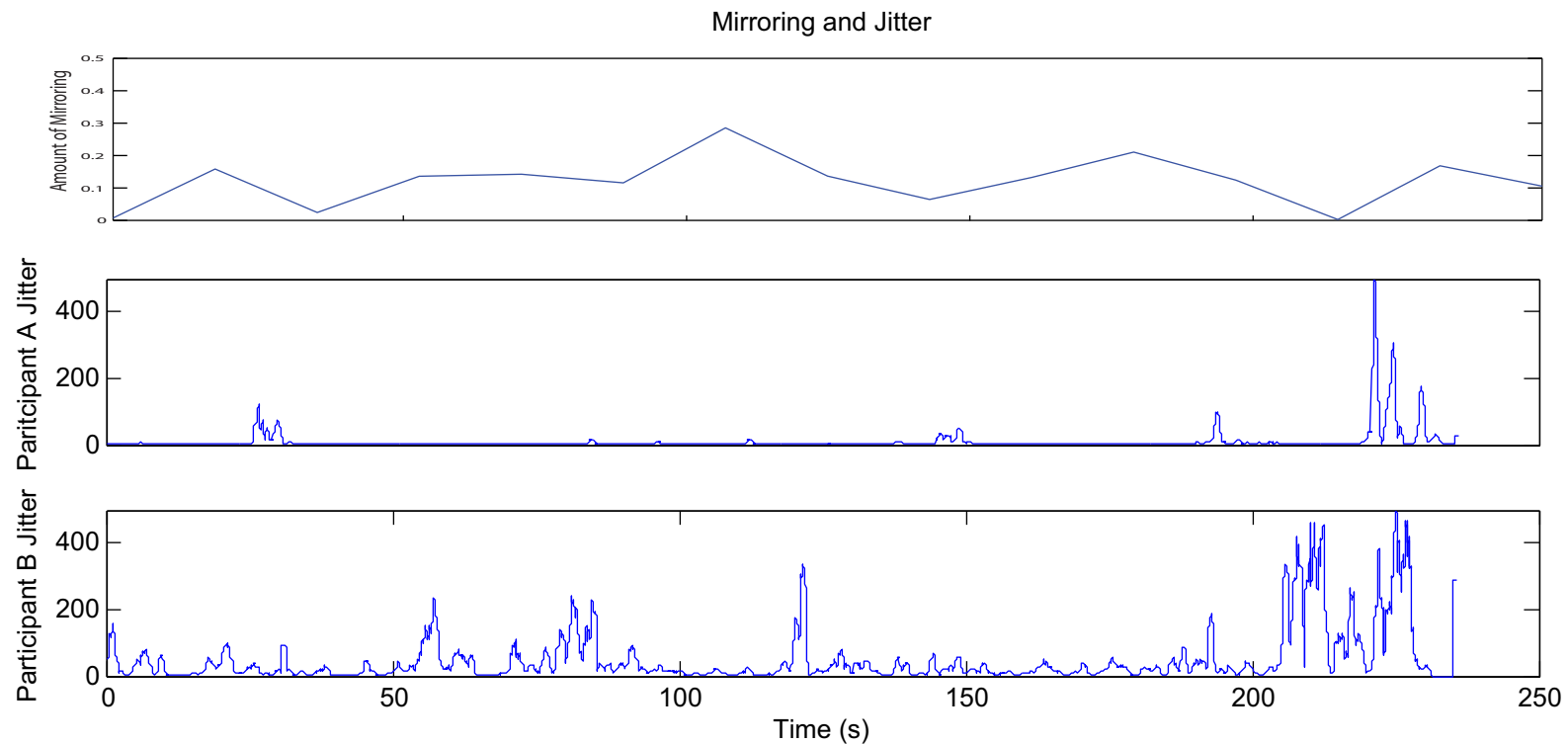


FIGURE 5-78 Cross-correlation and mirroring score between two people in a conversation, known to be disengaged and not mirroring until the end

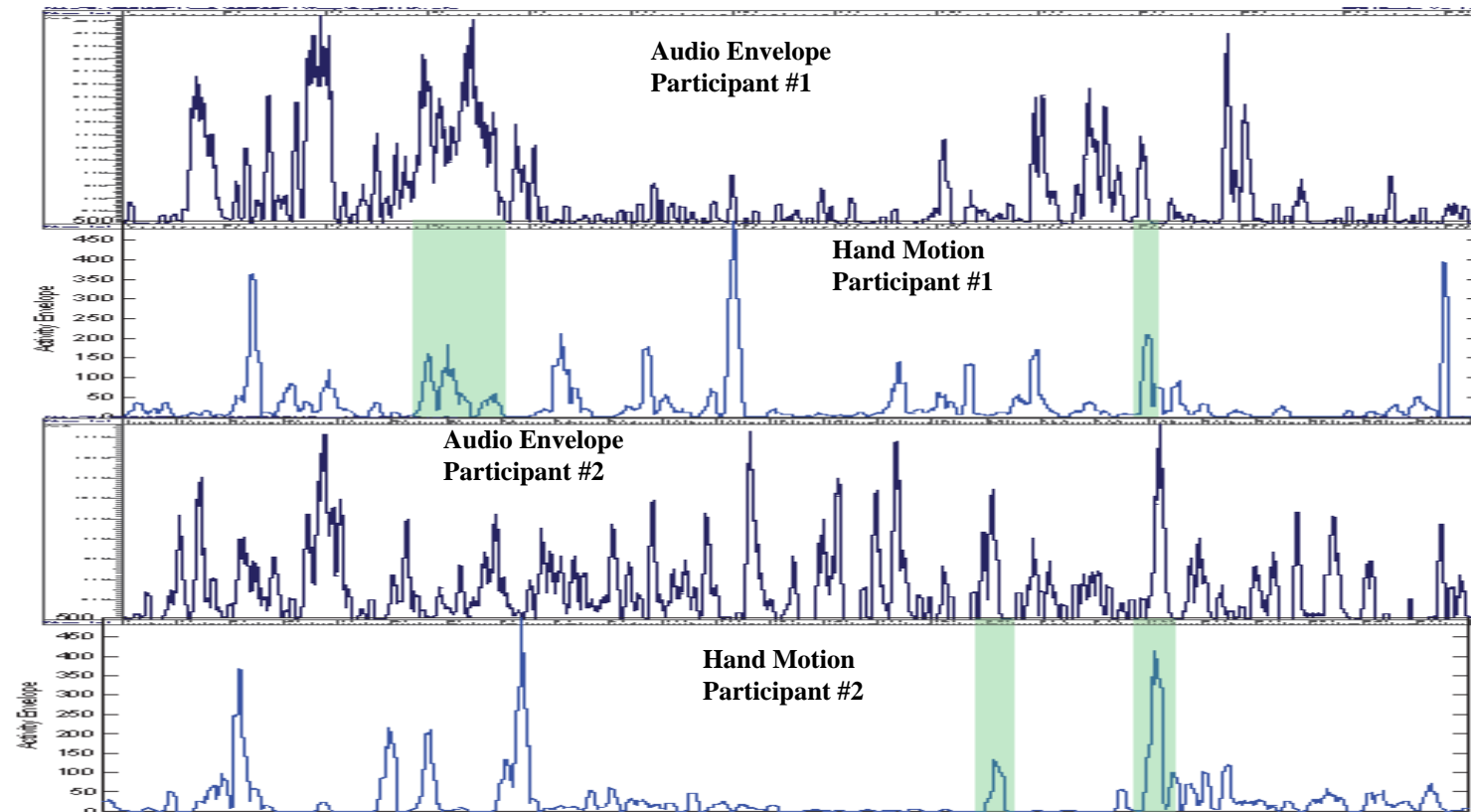


**FIGURE 5-79** The jitter is calculated from the body motion intensity while listening with the mirroring score subtracted. So the higher the mirroring the lower the jitter. Due to the large window required to calculate mirroring, there is a delay in the mirroring plot. Both the jitter value and the mirroring value while listening are used in the listener energy calculation



The speaker energy is based on the body language of emphasis. While speaking, the hand motion is a very accurate depiction of emphatic speech and is the major contributor to the speaker energy value. Unlike while listening, more hand motion while speaking indicates a higher level of energy. Posture and general body motion activity are also included, but at lower weights to the hand motion and arm angle.

**FIGURE 5-80** Hand activity while speaking is highlighted. The amount of hand motion delivered while speaking is a good indication of the amount of emphasis the speaker is putting behind his/her speech. The green areas on the plot are automatically created to show areas of emphatic gesture. The level of hand motion intensity while speaking is the main component of the speaking energy parameter



From the energy levels of speaking and listening from each person in a social interaction, the conversational energy can be calculated by the adding the zero-aligned speaking intensity and the listening intensity. The following figures show the speaking and listening energies of two participants, followed by the conversational energy of one of the participants as an example.

**FIGURE 5-81** Participant 1's Speech Intensity and Listening Intensity/Attentiveness, the value is set to zero for speaking intensity while listening and set to zero for listening intensity while speaking

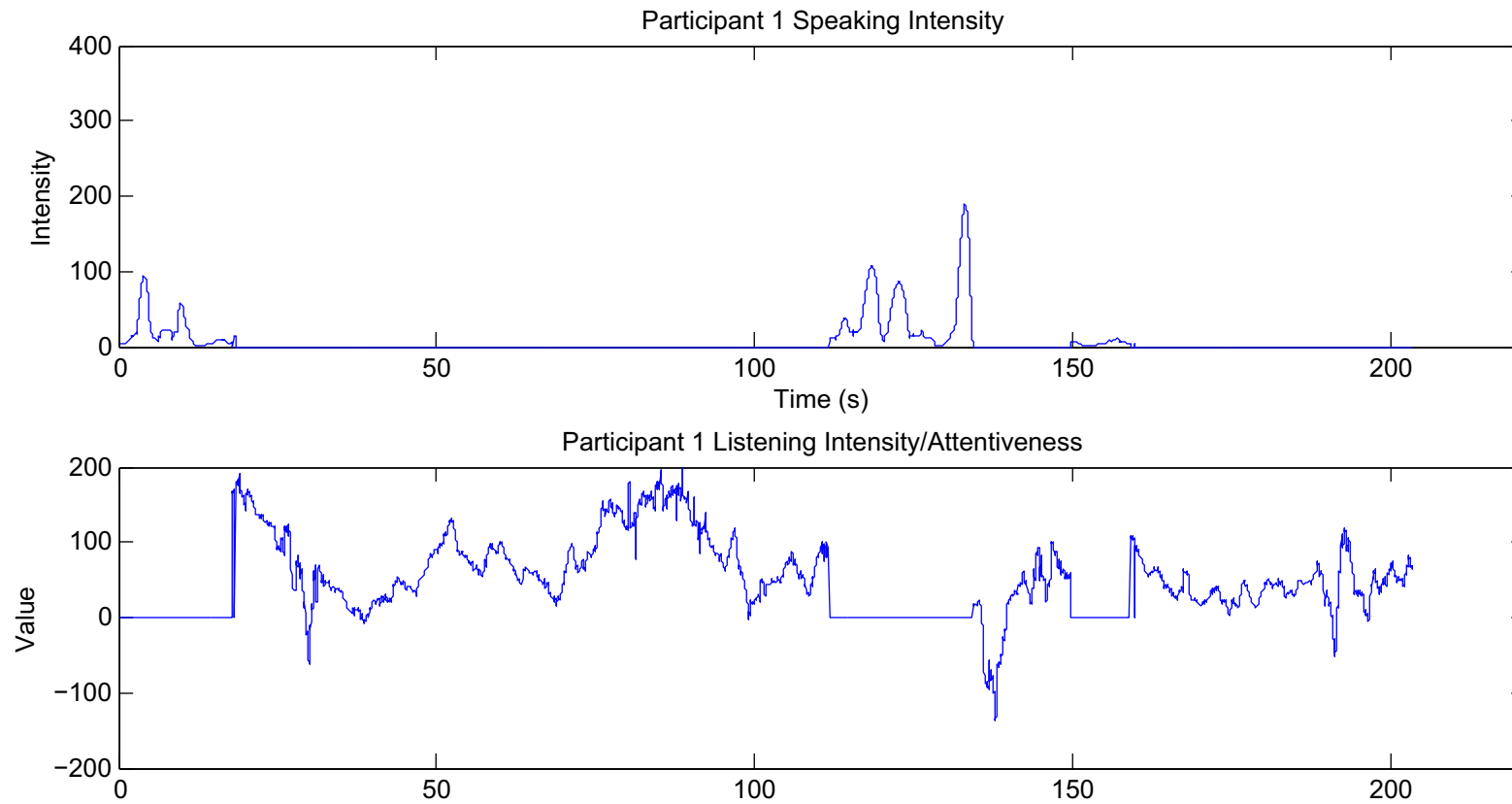
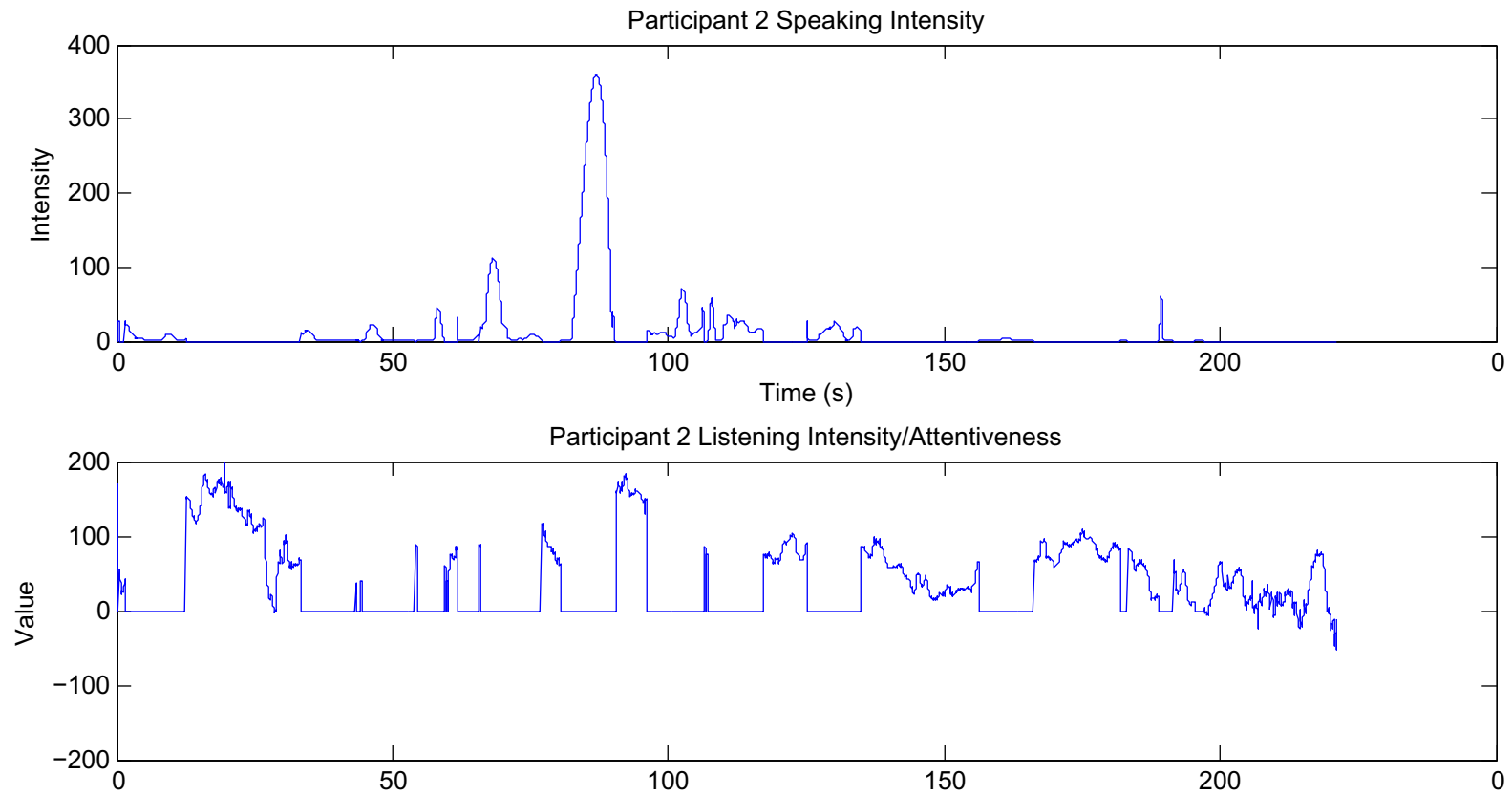
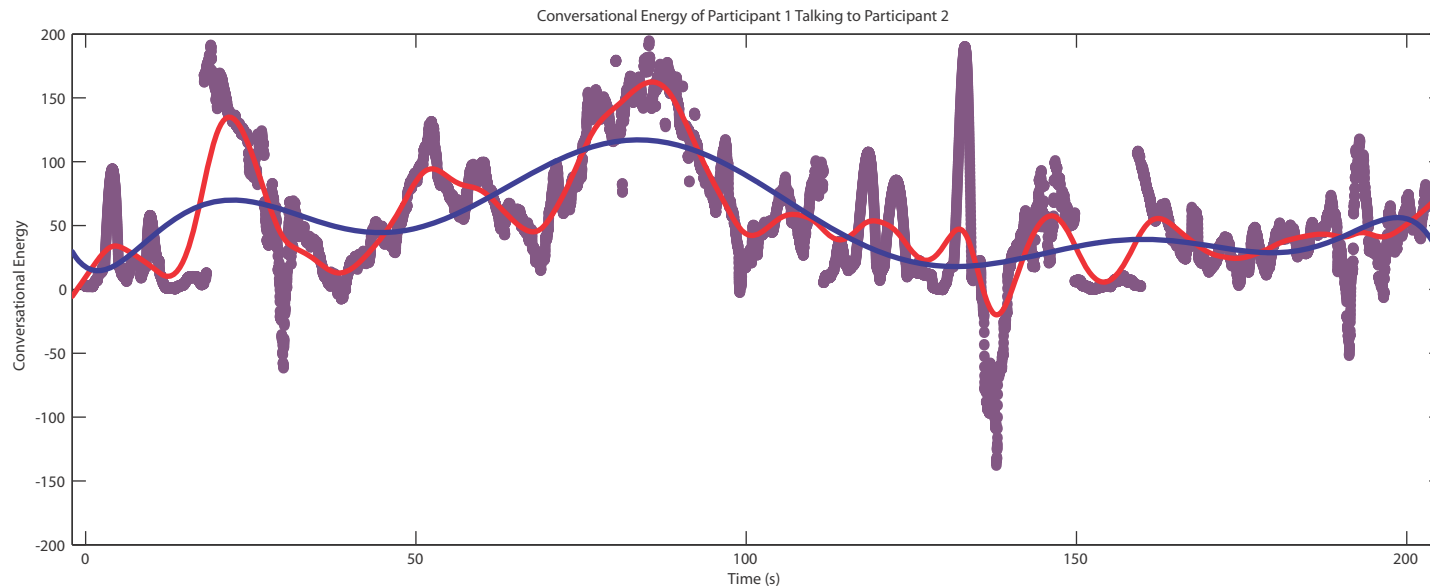


FIGURE 5-82 Participant 2's Speech Intensity and Listening Intensity





**FIGURE 5-83** The conversational energy of Participant 1 is calculated by adding the zero-aligned speaking and listening energy values



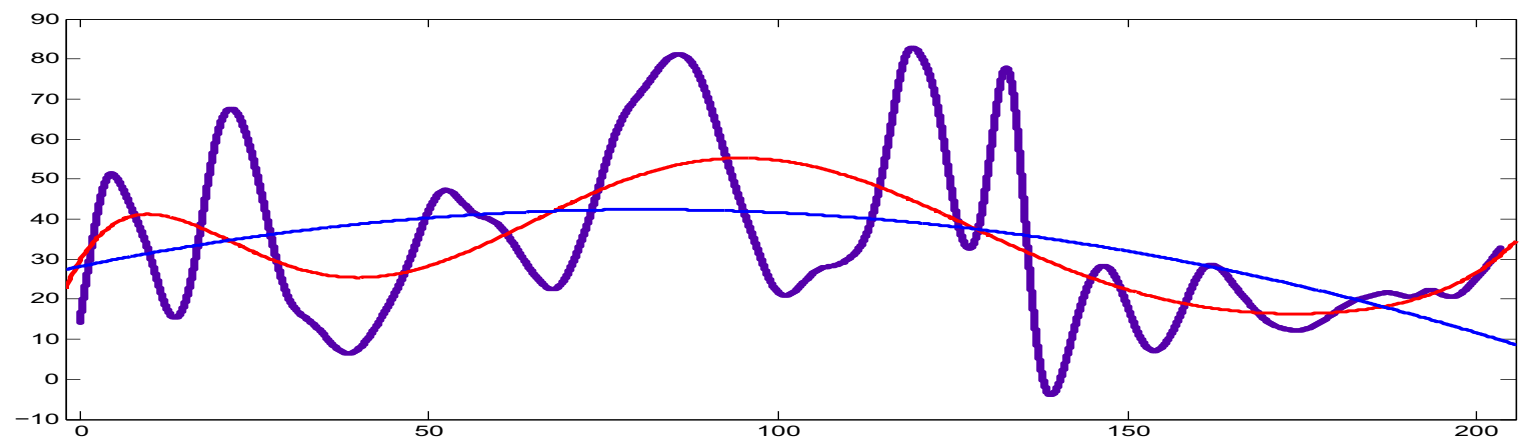
The above plot shows the energy expended by Participant 1 while talking to Participant 2. If you followed the plots of this section so far, you can see that at around 90 seconds into the conversation, Participant 2 speaks with increased volume and emphasis. At this time Participant 1 is seen to be listening quite intently as at this point is the peak mirroring, the lowest deviation angle, the local peak posture, and an absence of hand gestures. A review of the video by a group of non-participants confirms this moment as the peak of the conversation from Participant 1's perspective.

The conversational energy is directly mapped to the relationship score in one direction from Participant 1 to Participant 2. A similar curve can be calculated for the opposite direction. These two curves can be averaged

together to get a general indication of the strength of the relationship between Participant 1 and Participant 2.

Many of the fluctuations of the various envelopes used to calculate the relationship score are still present in the final values. This is exemplified by moments where no one is speaking or a lull between signals. Due to these fluctuations, the relationship score is interpolated to look at major trends as opposed to instantaneous data. The interpolation shown in blue in **FIGURE 5-84** has been hand-tweaked based on observation of the videos from various social interactions. How accurately the relationship score describes the social interaction is evaluated by comparing it to observed values as part of the survey described in Section 5.3.

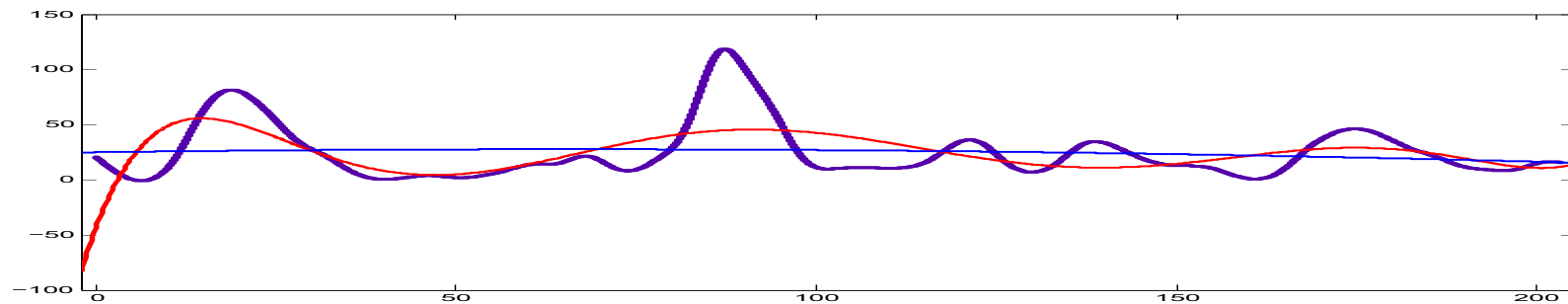
**FIGURE 5-84** The relationship score over the course of one conversation between Participant 1 and Participant 2. You can see the aforementioned spike at 90 seconds, but in general the conversation goes downhill from there and dwindles into a fairly indifferent interaction



The relationship curve shown in **FIGURE 5-84** actually represents one of the most dynamic relationships seen by the system to date, with a displacement of more than 20. When the conversation was observed through

video, almost everyone clearly saw the peak of interest followed by its dwindle. For comparison, the relationship curve from a conversation that matches almost exactly the average relationship score and variance from the entire bank of collected data is shown below.

**FIGURE 5-85 Example of the most common relationship curve, quite flat**



The above plot also illustrates the interpolation as the single major peak does little to affect the general trend. This is similar to how these events are viewed via the video, unless activity is sustained or of an amplitude greater than the one shown above, it will largely go unnoticed.

### 5.2.4 Speech Tonal Analysis

The badges can record high quality audio of the speech of the wearer. The system uses the speech envelope to determine when the person is speaking and when they are listening.

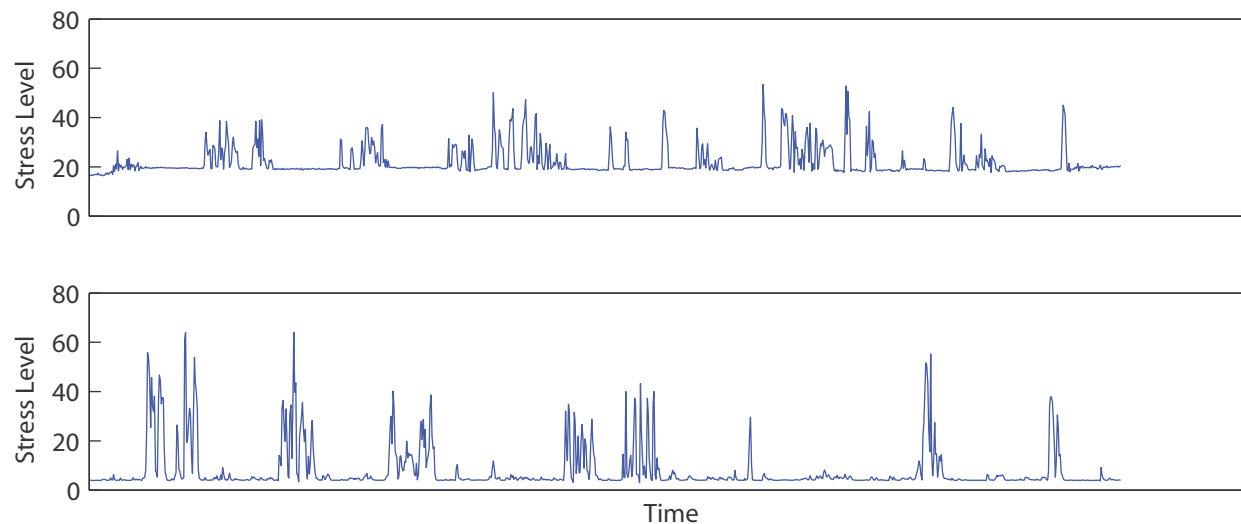
By recording the full audio wave, the system can pull out features by analyzing the waveform. Developing algorithms for the analysis of speech to understand affect is a major field of research.

To demonstrate the SPINNER platform's capabilities to support this field of research, a voice stress analysis routine was developed on the badge device. The routine is based on the McQuiston-Ford algorithm, which looks at the

energy in the frequency range of 8Hz to 12Hz. If the frequency with the peak amplitude between 8 and 12 hertz shifts quickly, this can indicate a tensioning of the laryngeal muscle, a sign a physiological stress. **FIGURE 5-86** shows the stress levels captured by the SPINNER badge during a conversation. Each peak is the stress level of a single utterance. Every utterance has some level of frequency shift, and this data should be viewed through a low-pass filter for dynamic qualities of stress.

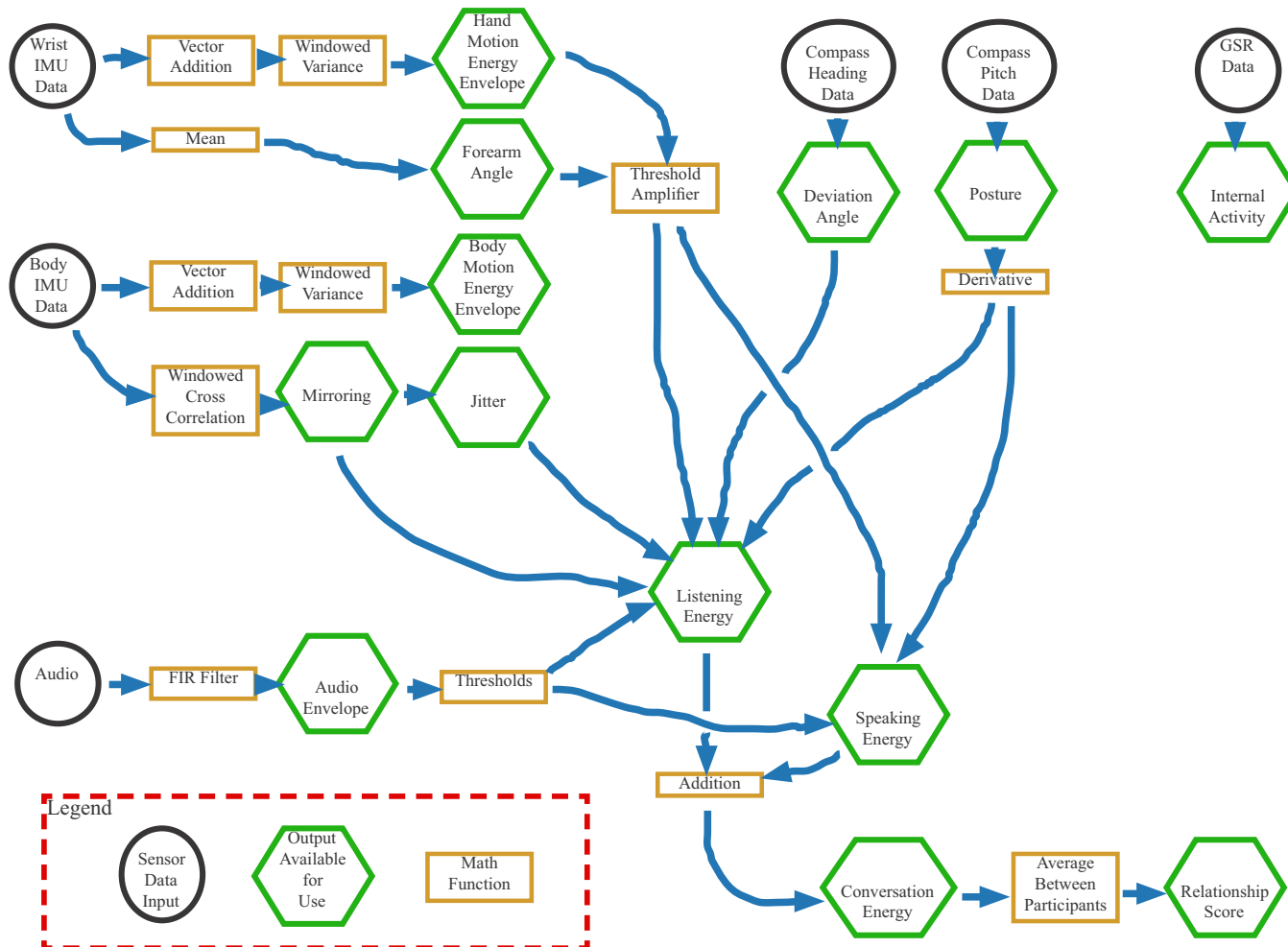
It is important to note that this is included as an example of the capabilities of the badge audio system. This has not undergone extensive testing or statistical analysis. This is a point of future work, but early tests have show promising results for future releases of the badge platform.

**FIGURE 5-86 Vocal Stress Analysis using Badges during a Conversation.** The participant whose stress analysis results are in the top plot was asked to try and stress the other participant out. When you remove the bias from Participant 1's data, you can see significantly higher peaks in Participant 2's stress analysis results



## 5.2.5 Summary Diagram of Sensor Data Processing

FIGURE 5-87 Block Diagram Summarizing the Processing from Raw Input to Outputs Available for Mapping to Story Energy and/or Labeling Video



|||||

**1. Is this a clip of:**

☐ A social interaction      ☐ Physical activity      ☐ Internal activity

**2. Please rate the level of activity (social, physical, or internal) based solely on observation and your general idea of the normal level of activity.**

	Zero activity (dead)	Minimal activity (sleep)									High level of activity	Extreme activity
Activity Level	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**3. Please write the moment in seconds from the start of the clip that you think the activity (social, physical, internal) level is:**

The Highest

The Lowest

**4. If this is a social interaction, please rate the interaction. If there are more than two people involved, rate the interaction between the two most active verbally. Consider Person A as the person you consider to spend more time on the left of Person B.**

	weak/disinterested/dislikes											powerful/attentive/likes
Person A's intensity while speaking	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person A's attentiveness while listening	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person B's intensity while speaking	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person B's attentiveness while listening	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

TABLE 5-5 Sensor Mapping Questionnaire (page 1 cont)

**4. If this is a social interaction, please rate the interaction. If there are more than two people involved, rate the interaction between the two most active verbally. Consider Person A as the person you consider to spend more time on the left of Person B.**

	weak/disinterested/dislikes										powerful/attentive/likes
Person A's intensity while speaking	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person A's attentiveness while listening	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person B's intensity while speaking	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person B's attentiveness while listening	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person A's feeling towards Person B	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Person B's feeling towards Person A	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**5. Does this clip provide you with any insight into the behavior of one or more people?**

	None whatsoever										I know them like I know myself
Insight	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

TABLE 5-6 Sensor Mapping Questionnaire (page 2)

**6. Which of these best describes the clip:**

☐ Something is lost

☐ Searching

☐ Finding

☐ Nothing of note

**7. Would you include this clip in a story about a day in the life of one or more of the people seen in the clip?**

Never  
include/pointless/nothing  
happens

Absolutely  
must  
include,  
clip tells  
all

Inclusion

☐☐☐☐☐☐☐☐☐☐

**8. If you rated 6 or higher to the previous question, please indicate in seconds from the start of the clip, the approximate moment that convinced/convinces you to include the clip.**

Time

The seven clips were run through the sensor data analysis to find the average relationship score and the peak activity value and point for a non-social clip. These values are then compared to the survey results. The remainder of the survey questions attempt to determine if the clip seems important as part of a larger story. Without a larger story, it is assumed that the apparent significance of a clip is related to how much change there is in the observable behavior of the characters. This is compared to the sensor data put through the basic story energy mapping.

The survey results are averaged from the 52 people who answered all the questions for all seven clips.

Creating Cohesive Video with the Narrative-Informed use of Ubiquitous Wearable and Imaging Sensor Networks 171



TABLE 5-7 Results from sensor data mapping evaluation

Clip	SPINNER social	Survey social	SPINNER peak	Survey peak	Survey/ SPINNER peak time error	SPINNER relationship	Survey relationship
Clip 11	Social	91.7% social	N/A	N/A	N/A	-2	45.
Clip 6	Not	60% not	81	74	1.3%	N/A	N/A
Clip 4	Not	100% not	50	53	5.6%	N/A	N/A
Clip 3	Social	90% social	N/A	N/A	N/A	45	84
Clip 9	Not	100% not	65	61	2.4%	N/A	N/A
Clip 12	Not	100% not	40	32	23.0%	N/A	N/A
Clip 13	Social	87.5% social	N/A	N/A	N/A	12	65

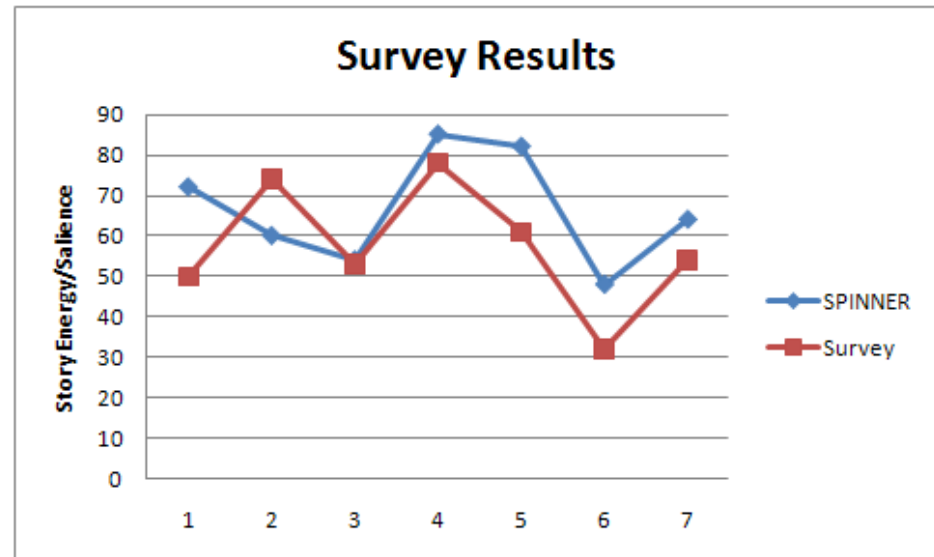
TABLE 5-8 Results from sensor data mapping evaluation (cont)

Clip	SPINNER story energy change	Survey clip salience
Clip 11	72	50.0
Clip 6	60	74.0
Clip 4	54	53.0
Clip 3	85	78.0
Clip 9	82	61.1
Clip 12	48	32.5
Clip 13	64	54.3

Although it is difficult to directly numerically compare the survey results to the sensor collected results, the relative numbers amongst the seven clips

can be correlated between the surveyed and calculated results. In general, the results track each other and show an approximate error rate of 18% across the entire table.

FIGURE 5-88 Charted Survey Results



Considering the potential for error in the survey results and the general tolerance of the application (the SPINNER system just needs to select and place the clips relative to each other), these results are decent enough to proceed with full-scale tests of the SPINNER Video Application. These tests are described with results in [Chapter 6](#).



# SPINNER VIDEO RESULTS

.....

.....

I meant exactly what I said: that we are saddled with a culture that hasn't advanced as far as science. Scientific man is already on the moon, and yet we are still living with the moral concepts of Homer.

- Michelangelo Antonioni[102]

The *SPINNER Video Results* chapter describes the actual use of the system to generate cohesive videos. It presents three specific executions of the system and evaluates their success.

## 6.1 APPLICATION SETUP

The SPINNER Video Application has been executed repeatedly while developing and testing the system. Each time the application has been executed a different set of inputs was used in order to create different videos. For each of these desired videos, a trajectory was designed along with specific mappings and edit points that effect the end result.

Three example videos have been selected, each with a different set of input parameters and a different level of complexity, and detailed below. Two of these videos were then evaluated for their content and effectiveness and compared to a randomly created video.

## 6.2 VIDEO #1 - MONTAGE

The first example video is actually the first successful video created by the SPINNER system. It was created without any wearable sensors to demonstrate the distributed video camera system, the narrative trajectory matching, and final assembly of a video.

Without the wearable sensors, story energy for each clip is detected by the sensors on the camera. It is mapped to the intensity of activity which is calculated from the motion sensor, the audio level, and the number of badges seen over the past window of time.

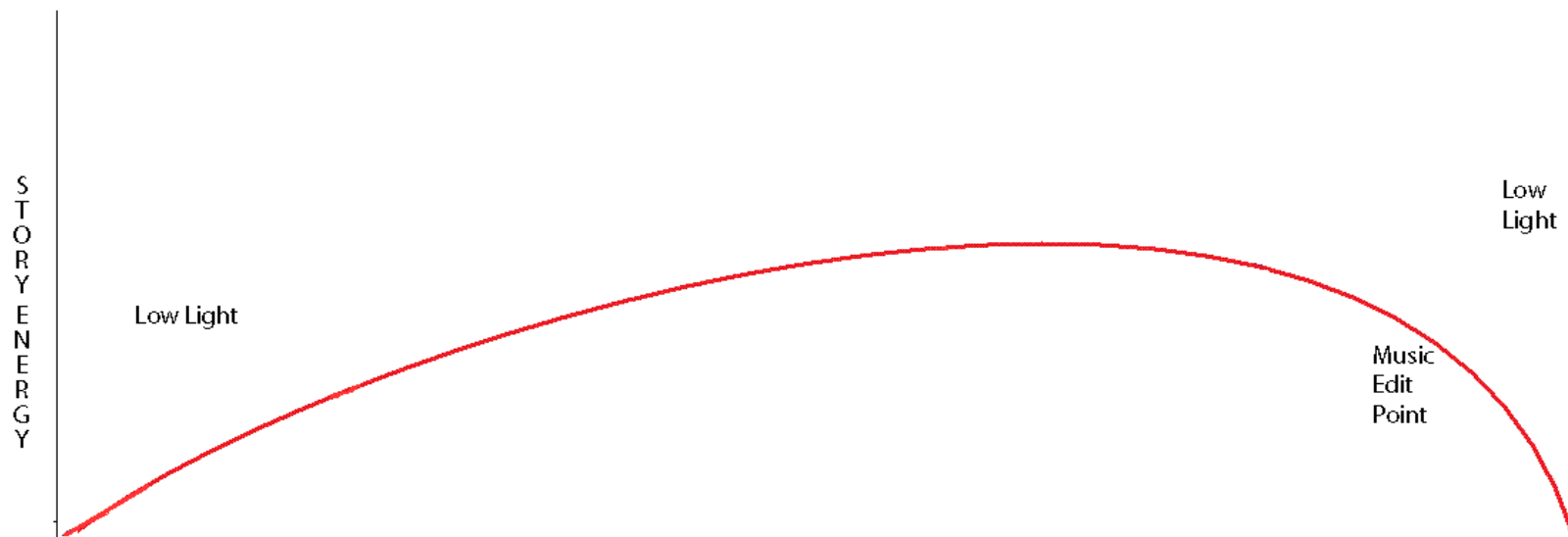
$$\text{activity} = 0.50\left(\frac{(\text{motion})}{100.0}\right) + 0.25\left(\frac{(\text{sound})}{32767.0}\right) + 0.25(\text{badges})$$

In the above equation, motion is defined as the percentage of time during the previous rolling window where the PIR sensor detected motion. Sound is defined as the amplitude of the sound pressure from the portal's microphone, and badges is either a one or a zero depending on if the portal has seen any badges at all.

In addition to the story energy trajectory for the video, the light level is mapped to the editing process. The clips selected in the first and last third of the final output are required to have lower light readings than the clips selected for the middle third.

The last input parameter is that an edit is required at 66 seconds in. This point was selected on the trajectory to be at the turn towards low energy.

FIGURE 6-89 Montage Video Trajectory



The video was created over the course of a two-day event with thirty-five active cameras resulting in over three weeks worth of raw video footage that the system needs to sort through. Upon viewing the video, those who were in attendance of the event said that the video accurately represented how they remember the event unfolding. Around 75% of those who were not in attendance said that the video made them wish they were there and gave them some insight as to what went on. The video is available for view on YouTube at [105].

**FIGURE 6-90** Montage Video Still



## 6.3 VIDEO #2 - JOURNEY

The second example video mainly illustrates the use of the wearable sensors to identify and locate the participants with respect to the cameras. This labels the clips with the subject. This video was created from close to two hours of raw footage collected from thirty-five cameras with seven participants wearing sensors, one of which was selected as the main character.

The trajectory designed for this video is completely flat. In other words, this video should be made from all clips that have a steady story energy curve. For this video, story energy was mapped to inertial motion variance during non-social clips. The system was instructed to only select from non-social clips, except for one social clip in the middle. The social clip selected should be the one with the flattest energy curve. For the case of the social clips, the energy curve for this video was mapped to listening energy. The general idea of these particular settings is to test the systems ability to track a main character and create a story based around slowly searching and finding one social moment to interrupt an otherwise steady adventure. This video is used as part of the evaluation study discussed in Section 6.5.

The audio in the Journey Video is mapped to the motion of the main character's wrist. The more wrist activity, the louder the music is mixed in with the sync sound recorded from the badge microphones. Since the trajectory is flat, there is not a lot of sustained wrist motion in the selected clips, but the effect is noticeable.



FIGURE 6-91 Journey Video Still



The pacing parameter was set so that the selected clips were not trimmed. The final journey video was around two minutes long edited down from over sixty hours of raw footage including one hour of raw footage that contained the main character. This video is available for view on YouTube at [106].

## 6.4 VIDEO #3 - STORY ARC

---

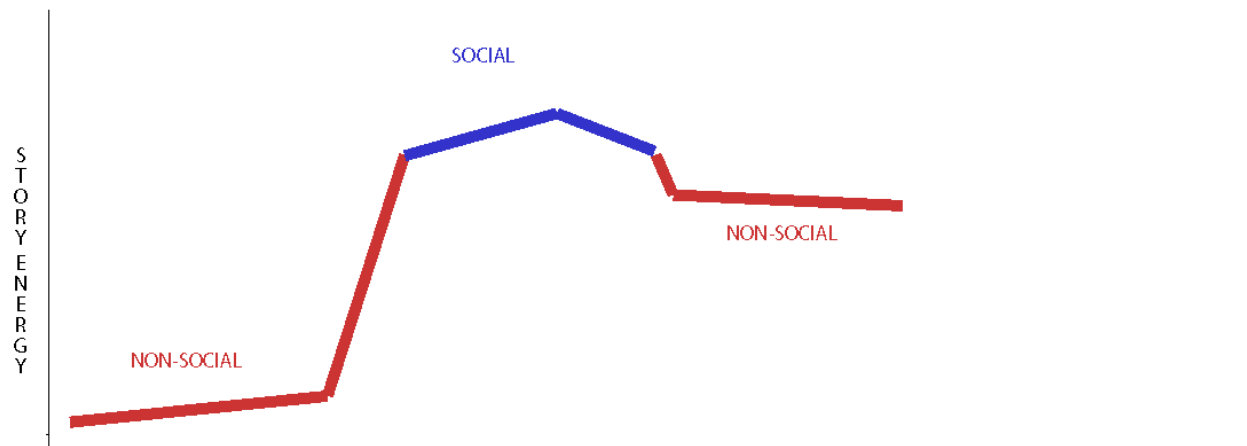
The final video example that was created by the SPINNER application illustrates the use of a standard story structure to sequence the clips. The

trajectory is based on the search model of narrative. This model is characterized by a period of slow activity followed by an up-tempo search. After the search is a period of experiencing the results of the search followed by a leveling off, but at a higher level than the start due to the experience of searching and finding. Countless stories can be fit to this structure.

For this specific video, we have set the story energy for clips showing a social interaction at a much higher baseline than clips that show the main character by themselves. This will then position social clips at the top of the trajectory, in the zone when it is considered that the main character has found what they are looking for.

The story energy value is created from wrist and body motion variance while alone, and a combination of speaking energy from the main character and listening energy from the characters that the main character is talking to while in a social interaction clip.

FIGURE 6-92 Story Arc Video Trajectory



The system is fed three tracks of audio and instructed to switch between the tracks every time there is a clip change from a non-social clip to a social clip and vice-versa.

**FIGURE 6-93** Story Arc Video Still, showing the main character at the height of searching. This is along the steep upward section of the story trajectory, just before he locates the social situation that he is searching for



This video was created during the exact same time frame as the journey video from the previous section. It too was created from over sixty hours of raw footage with seven participants including an hour of footage containing the main character. This video shows an entirely different perspective of events from the same time as the journey video. In each case, when the two main characters crossed paths, the specifics of the meeting did not meet the story conditions and the clip was not selected. This illustrates how the story trajectory can be used to completely separate series of events based on the narrative thread through actual time.

This video is used in the evaluation described in Section 6.5 where it is compared to the Trajectory Video and a video built randomly. The Story Arc Video can be viewed on YouTube at [107].

**FIGURE 6-94** Story Arc Video Still, showing the height of social activity at the peak of the trajectory curve



## 6.5 EVALUATION OF VIDEO RESULTS

In order to evaluate the SPINNER Video Application, the videos that it has created have been uploaded to YouTube so that they may be viewed and rated by a large audience. The Journey Video and the Story Arc Video as described above have been uploaded along with a video made by a random

assembly of clips also from the SPINNER cameras. The randomly assembled video is available at [108].

After viewing the three clips, the reviewers were asked to fill out the following questionnaire:

**TABLE 6-9 SPINNER Review Questions**

**1. Which of the three clips, if any, looks the most like it was made by a machine?**

- ☐ Movie 1
- ☐ Movie 2
- ☐ Movie 3
- ☐ None, they all look like they were created by humans

**2. Please rate the movies on how much you enjoyed or did not enjoy watching it.**

	Sucked										Awesome
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**3. Does the movie seem to show events in the order that they actually happened?**

	Completely Random										Precisely in the order that you think they would have occurred
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

TABLE 6-10 SPINNER Review Questions (cont)

**4. Do you feel that the movie provided you with any insight into the world of the characters?**

	Who were those guys again?									It's like they are in my head
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**5. How memorable is the movie? Feel free to come back and answer this later.**

	What movie?									You mean, that didn't actually happen?
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**6. Do you think the movie told a story? Use your own definition of what you think it means to tell a story.**

	It was like Avatar without the special effects	It was like Avatar even with the special effects								Coppola is a hack comparatively
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**7. The movie's length was:**

	Too short				Perfect				Too Long
Movie 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Movie 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sixty-seven people took the survey. The average results are shown here.

**TABLE 6-11 SPINNER Review Survey Results**

Question	Travel Movie	Story Arc Movie	Random Crap
Enjoyment (1-10)	5.83	7.02	4.01
Memorable (1-10)	5.58	6.5	4.08
Insight (1-10)	5.58	6.42	3.08
Told a story (1-10)	5.18	6.09	3.09
Length (5 is perfect)	4.67	4.58	3.42
Mechanized feel (%)	0%	18.2%	82.8%

The results show that both the Story Arc Movie and the Travel Movie were enjoyed substantially more than a randomly assembled video, despite all three movies coming from the same source material. Almost everyone selected the random movie as the one most likely assembled by a mechanized process. In general, the results show that the videos created with by the narrative-informed SPINNER system were enjoyed in general as well as in comparison to the video edited by a process without structure. The one strange result is that the random video that people did not enjoy that much was seen as too short. This is probably due to the fact that it has no meaning and it is assumed that something is missing that would be included if it were longer. The other two

videos are considered as near perfect length, indicating that they are viewed as being a complete story with nothing missing.

These results, combined with the fact that it would take a human hundreds of hours to filter through the video and recreate this process, tell us that these techniques are promising and the SPINNER Video Application is a shining success.





# EPILOGUE

---

A tragedy is a representation of an action that is whole and complete and of a certain magnitude. A whole is what has a beginning and middle and end.

- Aristotle [104]

The *Epilogue* chapter, commonly called the conclusions, summarizes the overall contributions of the work and presents some ideas of what could be done next now that this research field has been created and proven as valuable.

## 7.1 CONTRIBUTIONS

---

Media technology has exploded into the very fabric of our lives, bringing with it a host of potential benefits and drawbacks. This dissertation can be seen as a call-to-arms for research in the area of ubiquitous media and experiential video content creation, without which there will be no way to understand and confront the drawbacks and allow the benefits to have their full impact on society.

The SPINNER media platform launches us head-first into the type of research that is necessary in the age of ubiquitous media. In addition to specifying and discussing the various attributes and tenants of this new era, this dissertation has developed and demonstrated an actual application that highlights what can be done as a result of this studied investigation. More specifically, this dissertation has contributed the following to the world:

- Powerful new media platform - The SPINNER Media Platform has provided a deployed and commodified system for all to experiment with distributed cameras, wearable sensing, social signal analysis, situated media technologies such as screens, wearable distributed audio capture, and the use of narratology to design applications and systems
- New form of entertainment and communication - The SPINNER applications allow you to create comprehensible video content out of your social and personal behavior. This method of creation promotes validity of output and puts the ability to create meaningful content into everyone's hands, in other words, not special knowledge or practice is requirement to create, it comes from one's actual experience

- Toolkit for documentation of daily life and self-reflection - The videos created by the SPINNER system provide a new perspective into life that can lead to new and unexpected insights about events that might go unnoticed while living them
- Discourse on narratology - Narrative is a means by which we understand events and information. This dissertation has shown the power of using this method of thought in unorthodox ways applicable to many projects and in many aspects of life
- Multi-tiered system architecture - The SPINNER system combines battery-powered mobile devices, situated camera devices, backend services, and purpose-specific interfaces always allowing the best access to phenomena and best use of resources. The effectiveness of this type of architecture was demonstrated and will provide inspiration for many systems to come
- Management of collected media - The SPINNER system uses wearable sensors, situated sensors, and other techniques for the real-time labeling of collected video, audio, and sensor information. This allows the collected and created content to be managed, organized, and ultimately browsed.
- Transforming the environment - The SPINNER system transforms the very environment in which we live and work into a creative tool. It also serves as a platform for new experiences within our everyday lives
- Deployment experience - There were significant challenges faced in order to deploy the system into the environment. Concerns with privacy and security were constant, leading to difficulty deploying the system in certain areas. In addition, people often did not behave naturally near the cameras and sensors. These issues will be faced

again and again in the future when trying to enhance the space with pervasive sensing systems. By addressing these concerns, collecting actual information about privacy and acceptance, and by providing benefits from having such a system, we have made a major step forward in our practical ability to perform research in this field.

## 7.2 FUTURE WORK

---

The overall success of the research described in this dissertation has inspired much thought as to where it should go next.

Each time the entire system has been run, significant amounts of data have been collected leading to iterations on the design and major insights into the worlds of ubiquitous media and content creation. Although the situated distributed camera/sensor network is always active, the main point of future work is to have the wearable sensors running constantly. This would collect endless amounts of data to develop new algorithms and applications on, and provide the most honest data as the participants become accustomed to the system. In short, the most important thing for the system in the future, is to keep using it.

To this end, work can be performed to allow the wearable sensors to be accepted and always on. This can be achieved by integrating the sensors into objects such as wrist-watches and cellphones. The addition of a first-person perspective camera to the system that could come from integrating mobile phones or synchronized wearable cameras would further the detail of the

output video. This could also fill in the gaps of coverage of the environmentally placed cameras. Adding more cameras, in general, would increase coverage and decrease the missed action.

There are, of course, endless technical upgrades that can be made to the system such as upgrading all the cameras to high definition. The software can be upgraded to include more mapping possibilities such as mapping the video transitions to the sensor data and generating music from the collected data and final edits.

Through continued experimentation with the SPINNER system, better definition and evaluation of the more elusive and subjective features relevant to social narrative can be achieved.

In addition to its use for user-created video content and life documentation, the SPINNER system is ideal for creating certain forms of new and traditional media. The most obvious application is for the ever-increasingly popular reality television market. This system could be used to automatically capture events from the set of a reality tv show, or allow a user to choose what types of events they would like to see.

Techniques developed in this system can be used to create an application that automatically annotates collected video to ease the job of a human editor or to automate certain editing tasks, even on a scripted shoot with many cameras and takes. This brings about the idea of the augmented living studio space, where a film can be made just from the actors acting in the space naturally. The scalability of the system would allow the studio to be extended indefinitely and the performance or shoot could travel throughout the reach of the device network and the captured video would still be synchronized and annotated.

And finally, the SPINNER system can be used with actors performing a specific story as opposed to real life. The video can then be projected onto

several different user-selected story curves, resulting in a new form of interactive cinema

## 7.3 FINAL THOUGHTS

FIGURE 7-95 **Keep on SPINNING!**



# APPENDICES

.....

.....

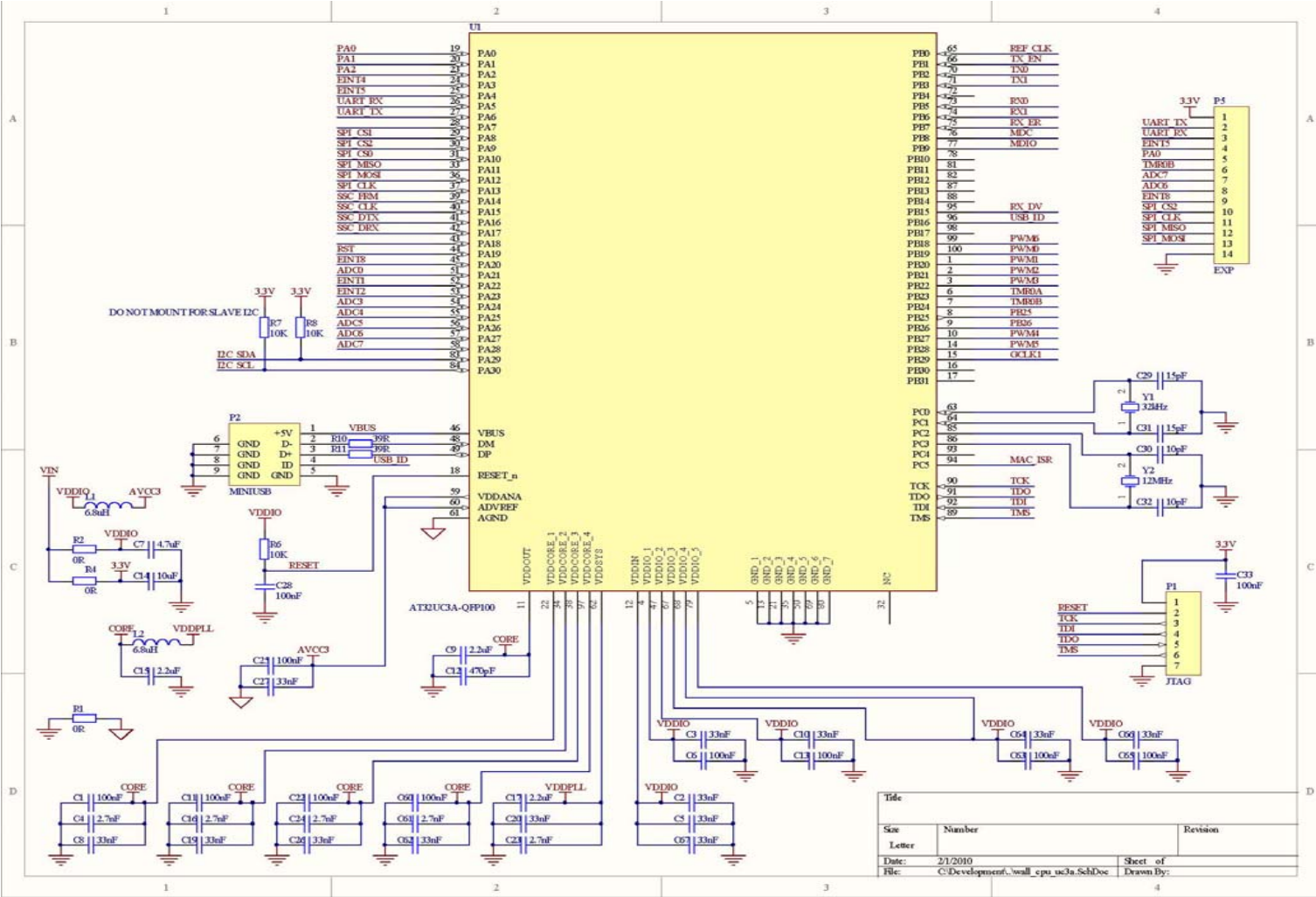
**Y**ou better memorize this face  
you better stay right in your place  
i draw the lines here from now on  
and your picture's already drawn  
and this movie goes on to long  
and this coffee's a little to strong  
and i think that i'm running on  
well i guess that i'm running

- Superchunk [109]

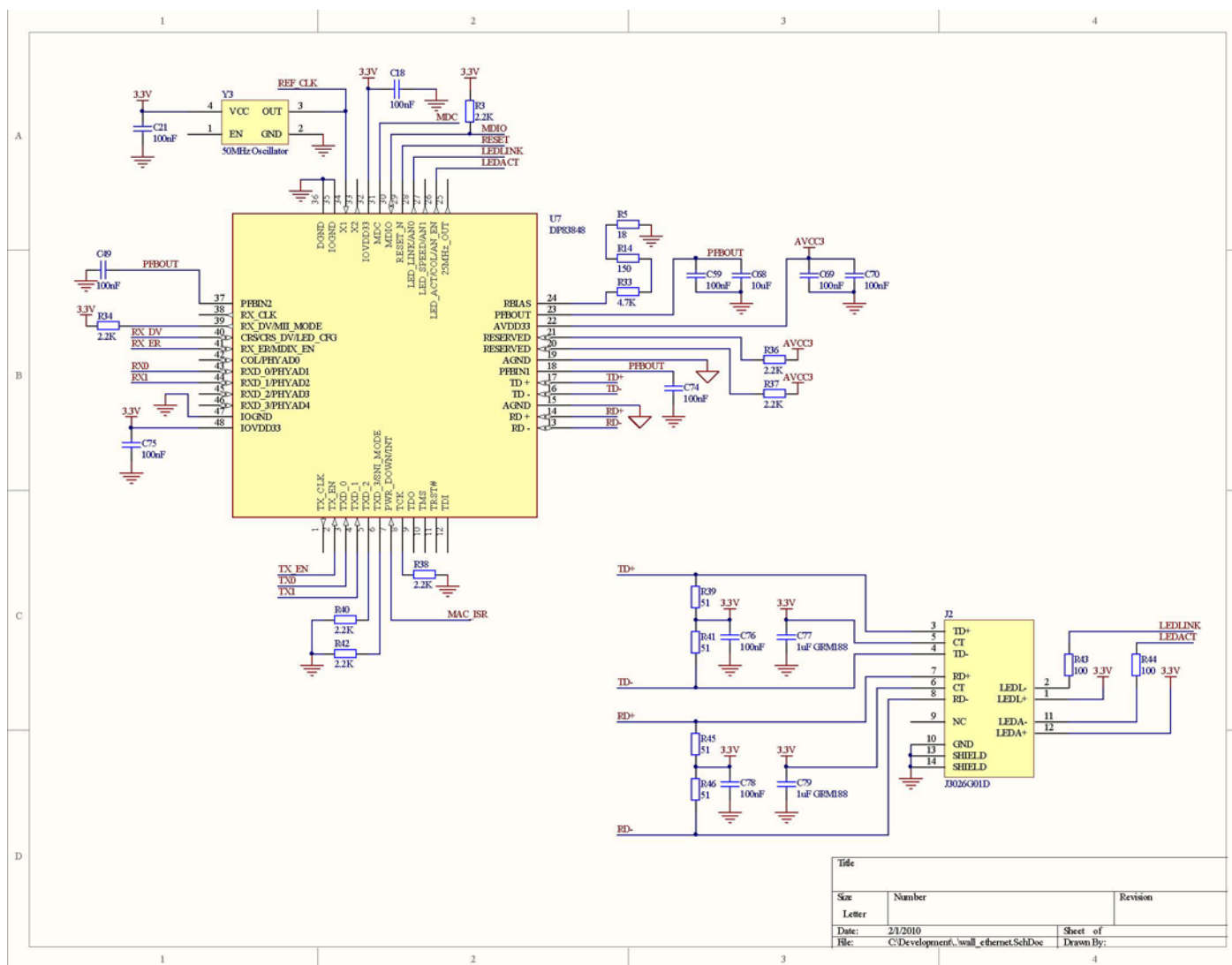
The Appendices chapter contains a sampling of the immense amount of technical documentation from the world of SPINNER.



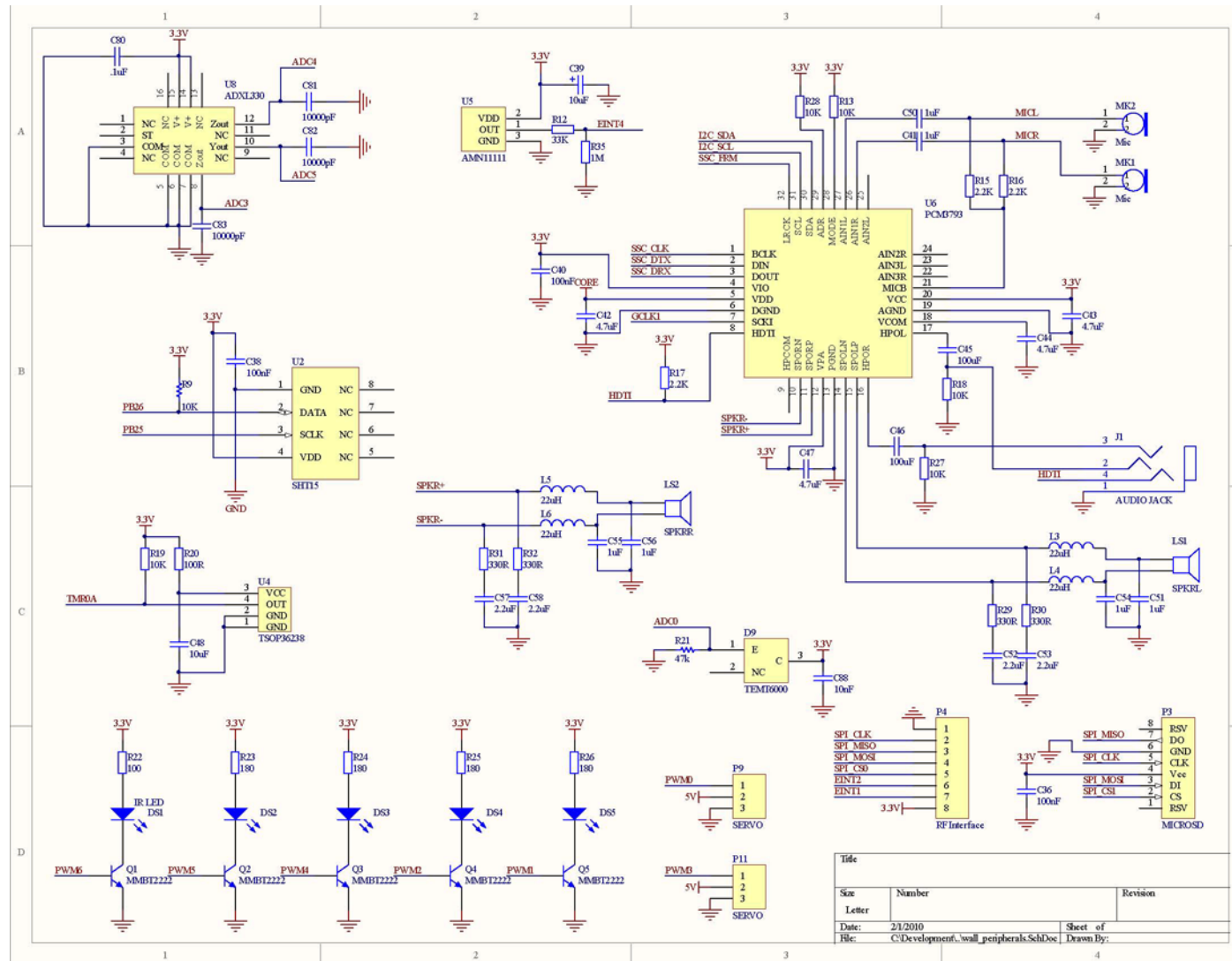
|||||



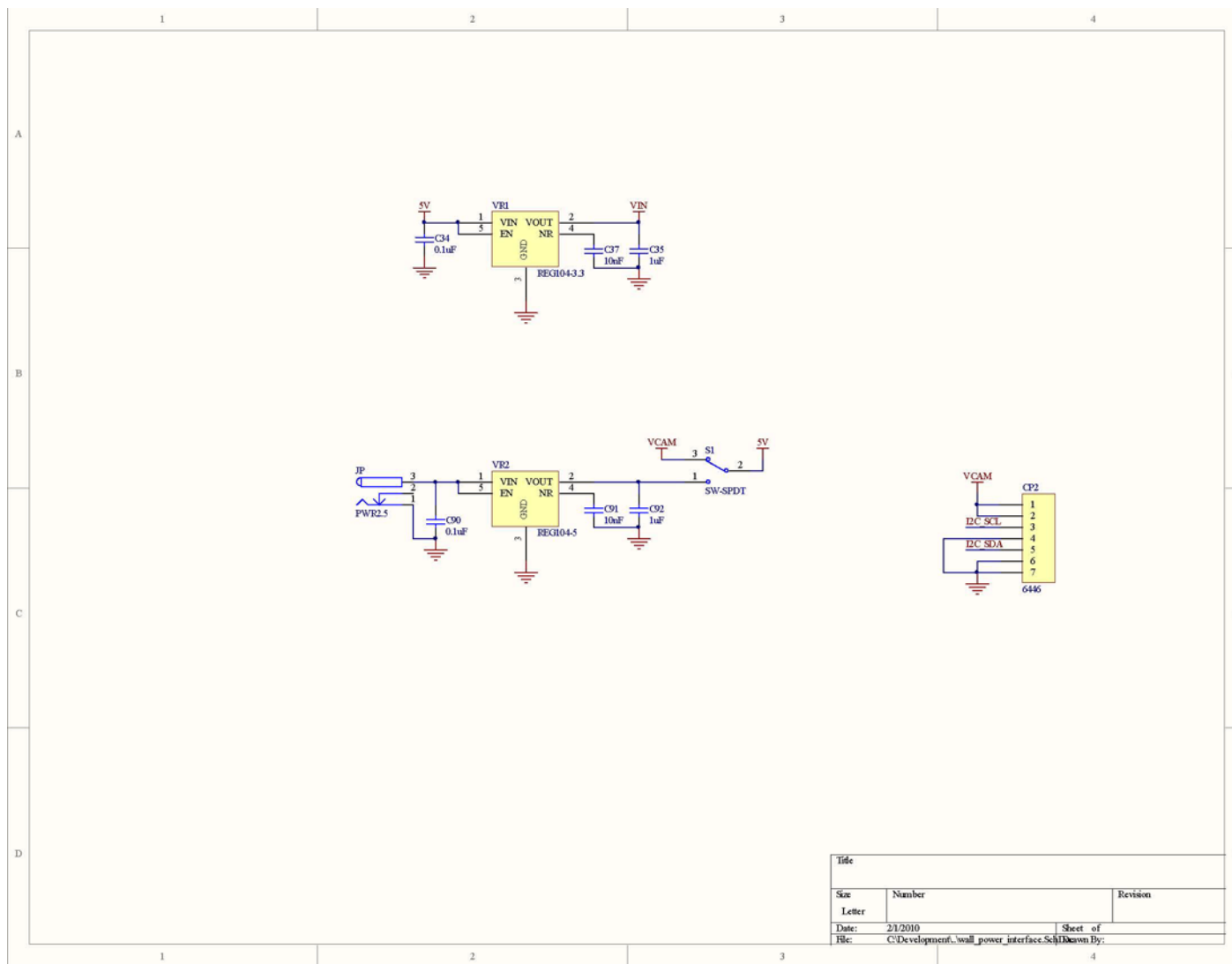
## A.1.2 Red Board - Ethernet



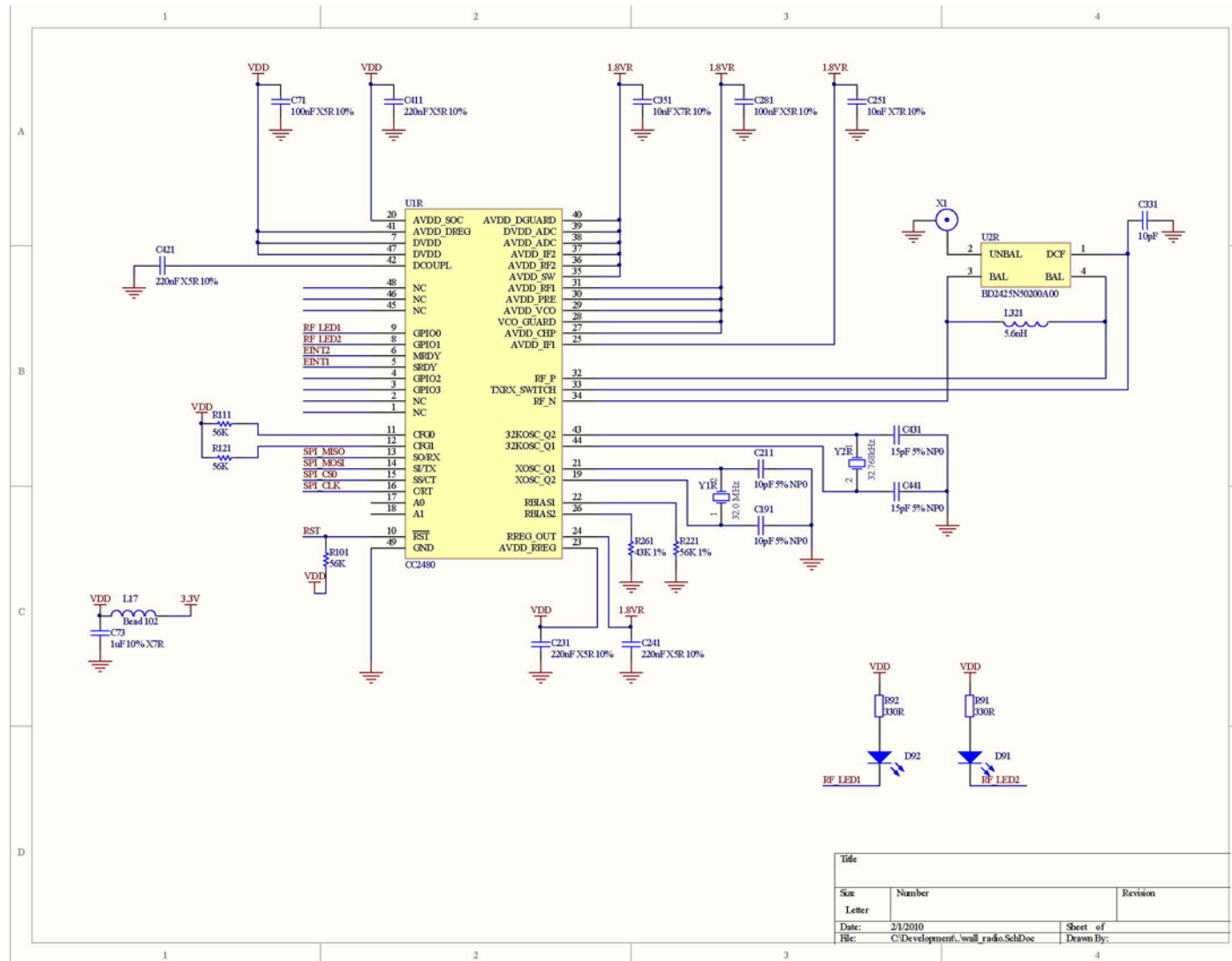
## A.1.3 Red Board - Peripherals



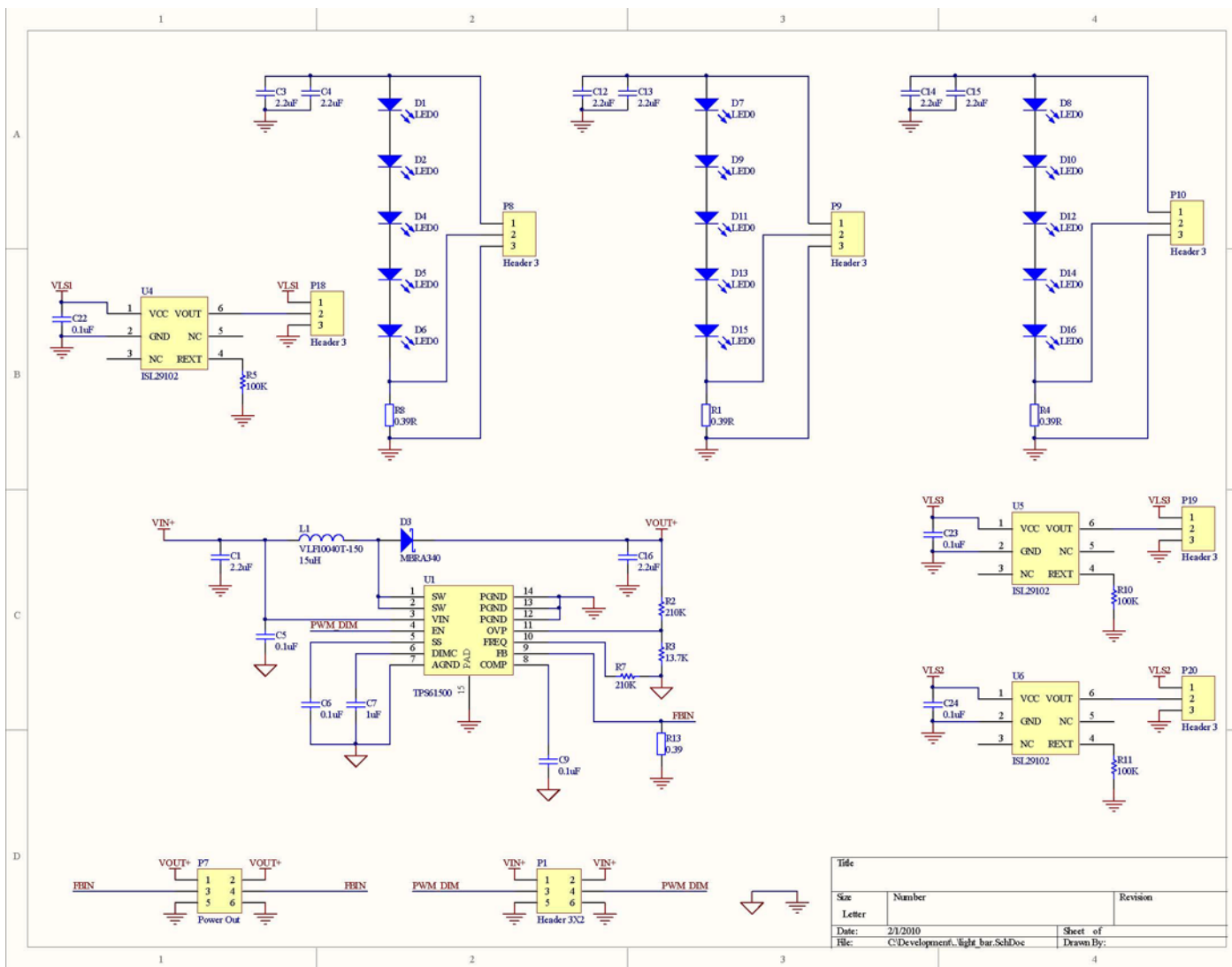
## A.1.4 Red Board - Power



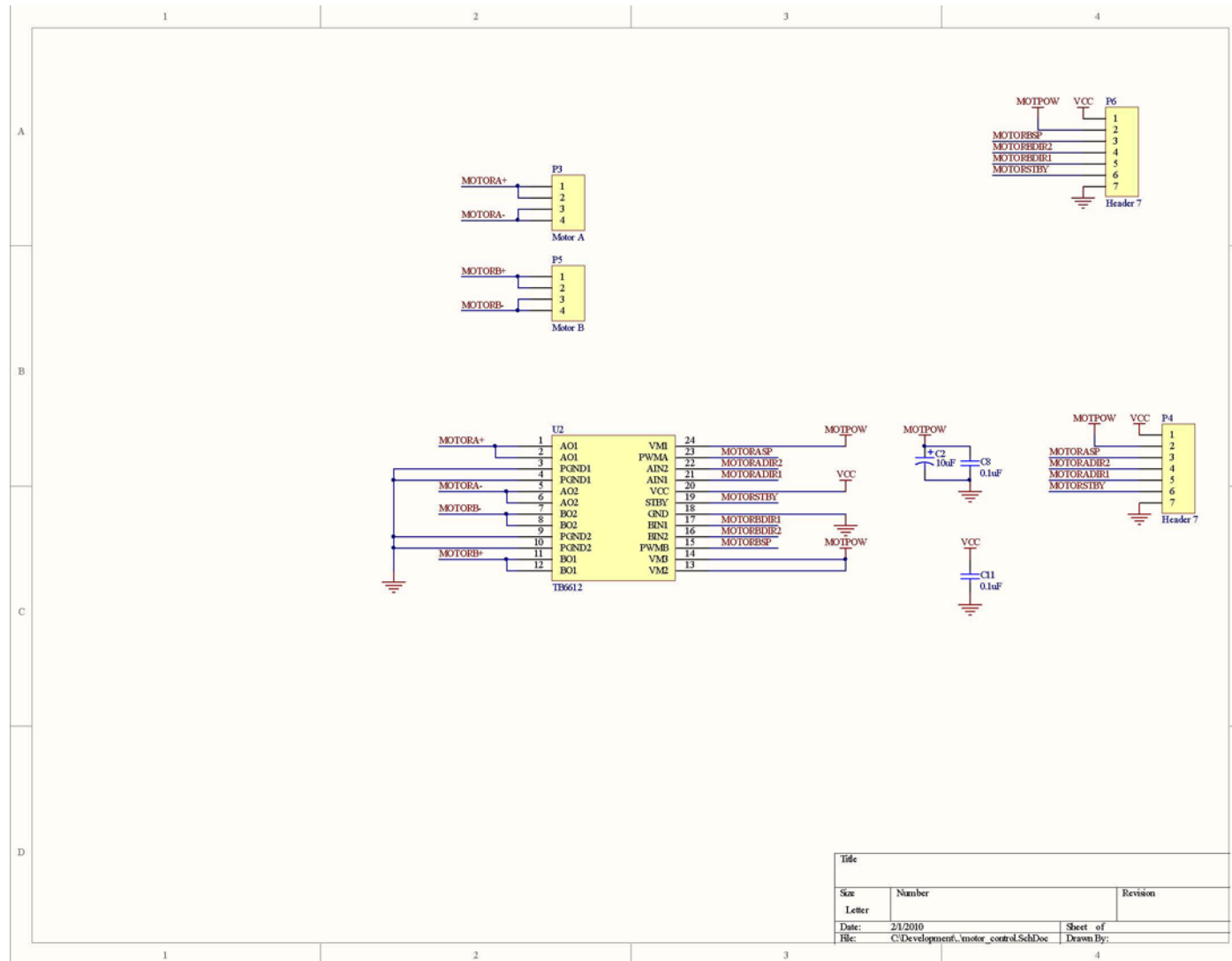
## A.1.5 Red Board - Radio



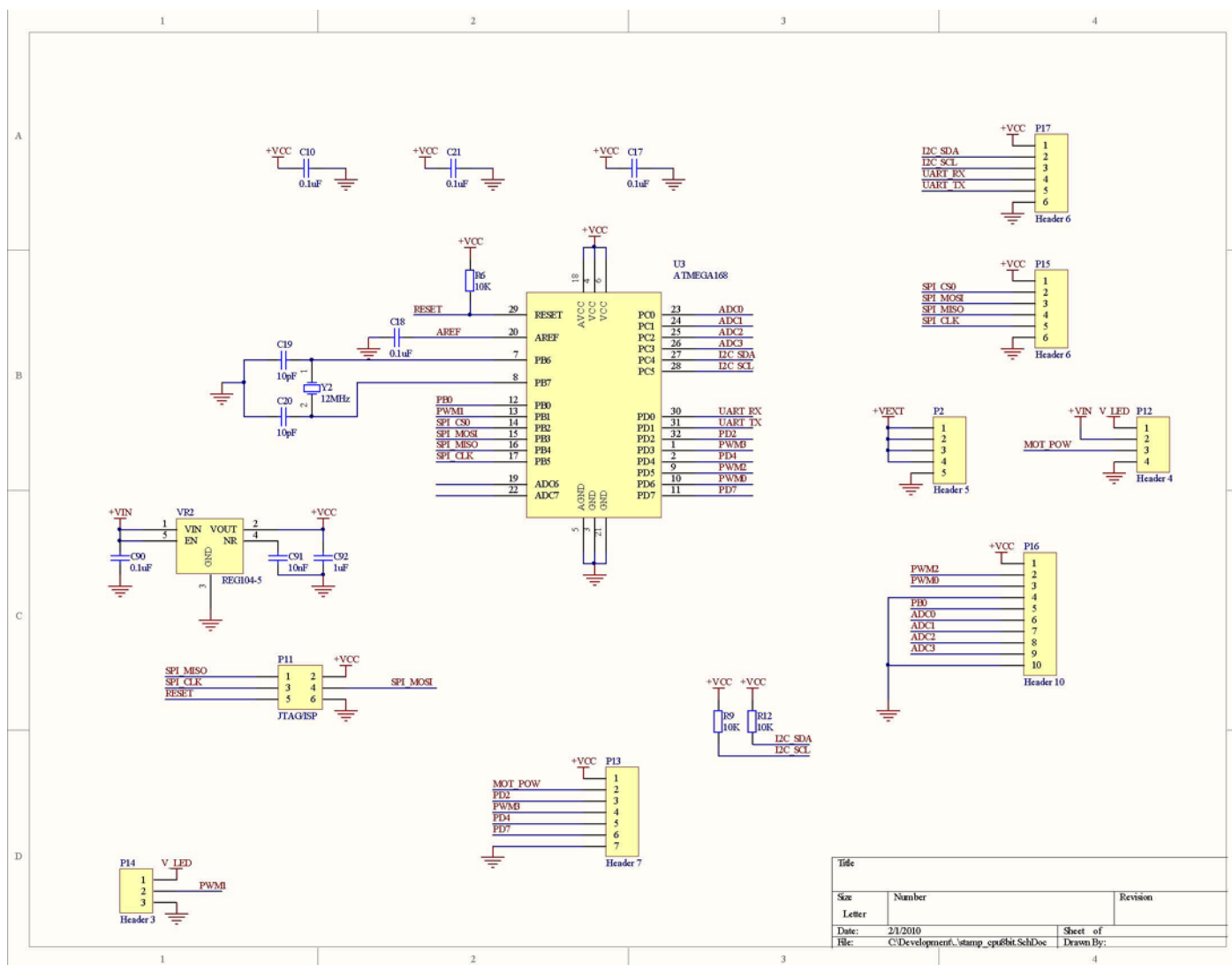
## A.1.6 Hat - LEDs and Controller



### A.1.7 Hat - Motor Controller

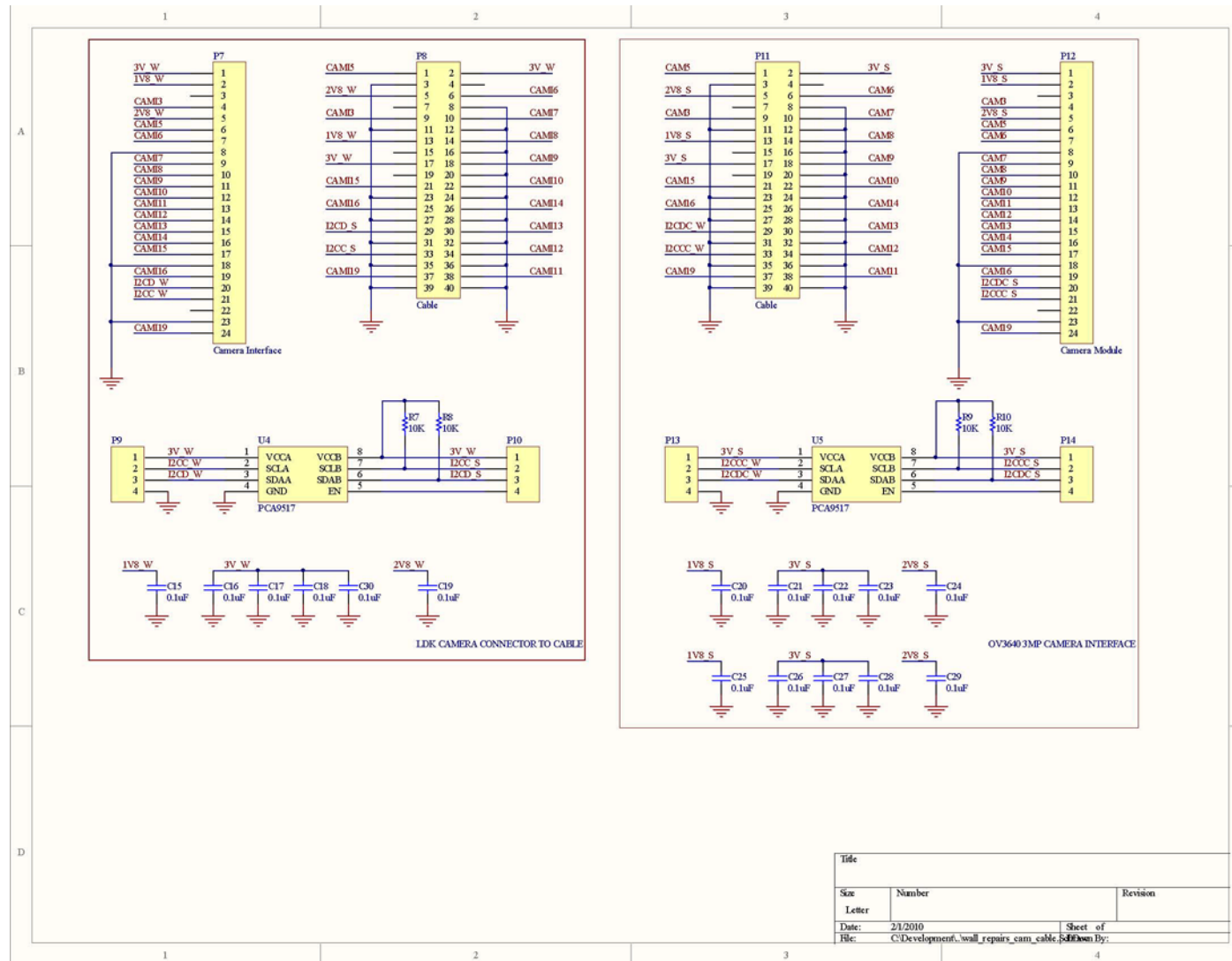


## A.1.8 Hat - Microcontroller

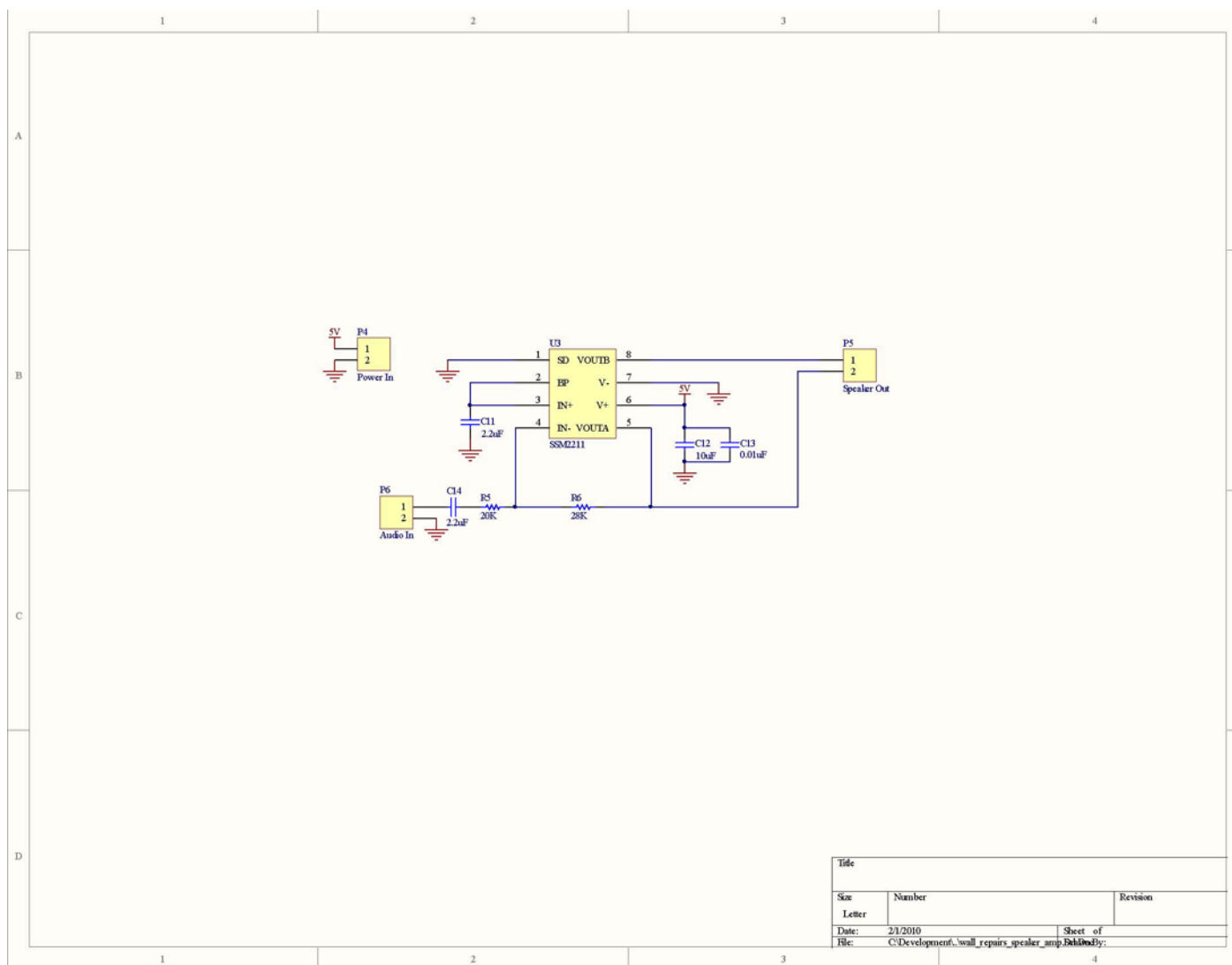




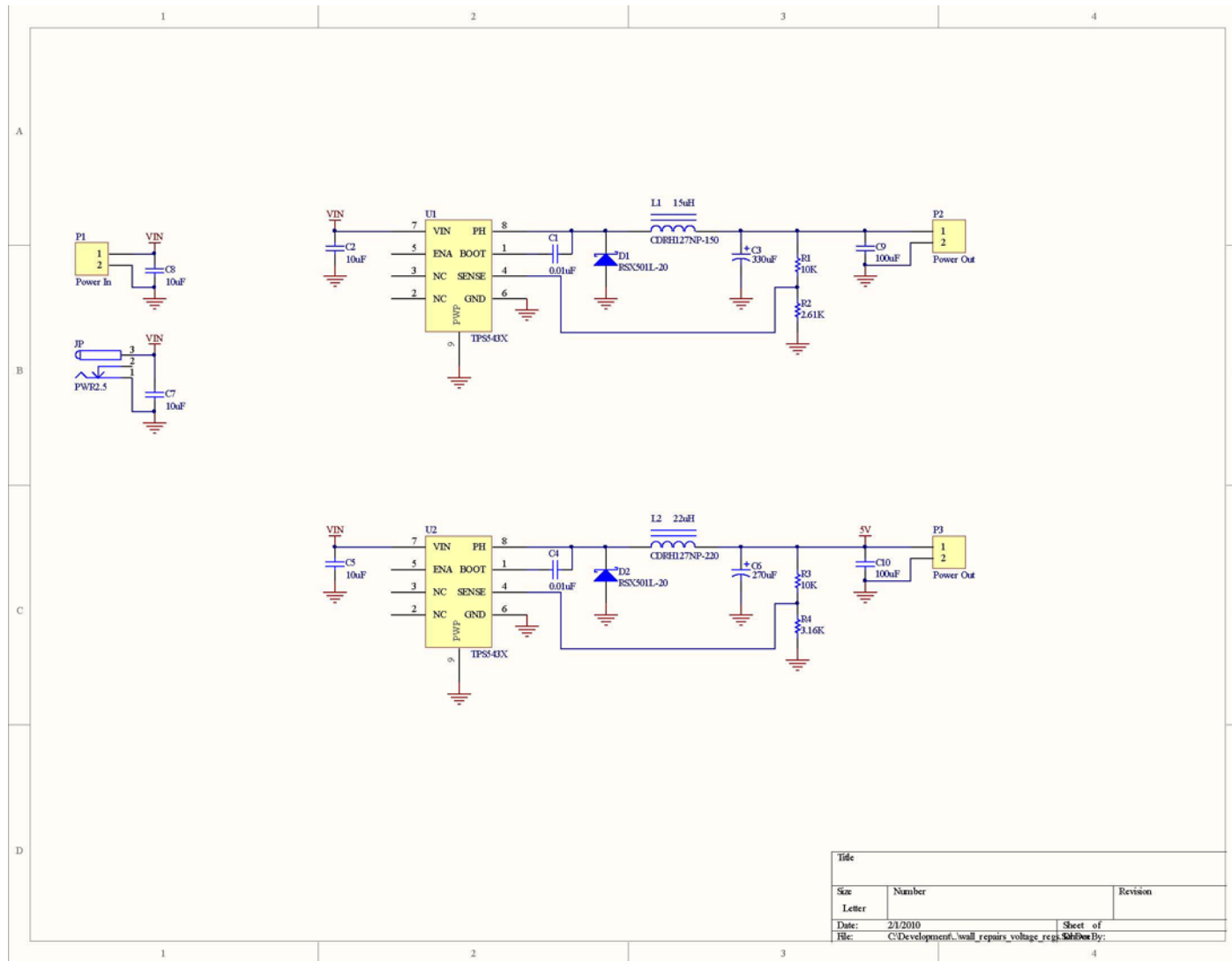
### A.1.9 Camera Cable and Buffers



## A.1.10 Speaker Amplifier



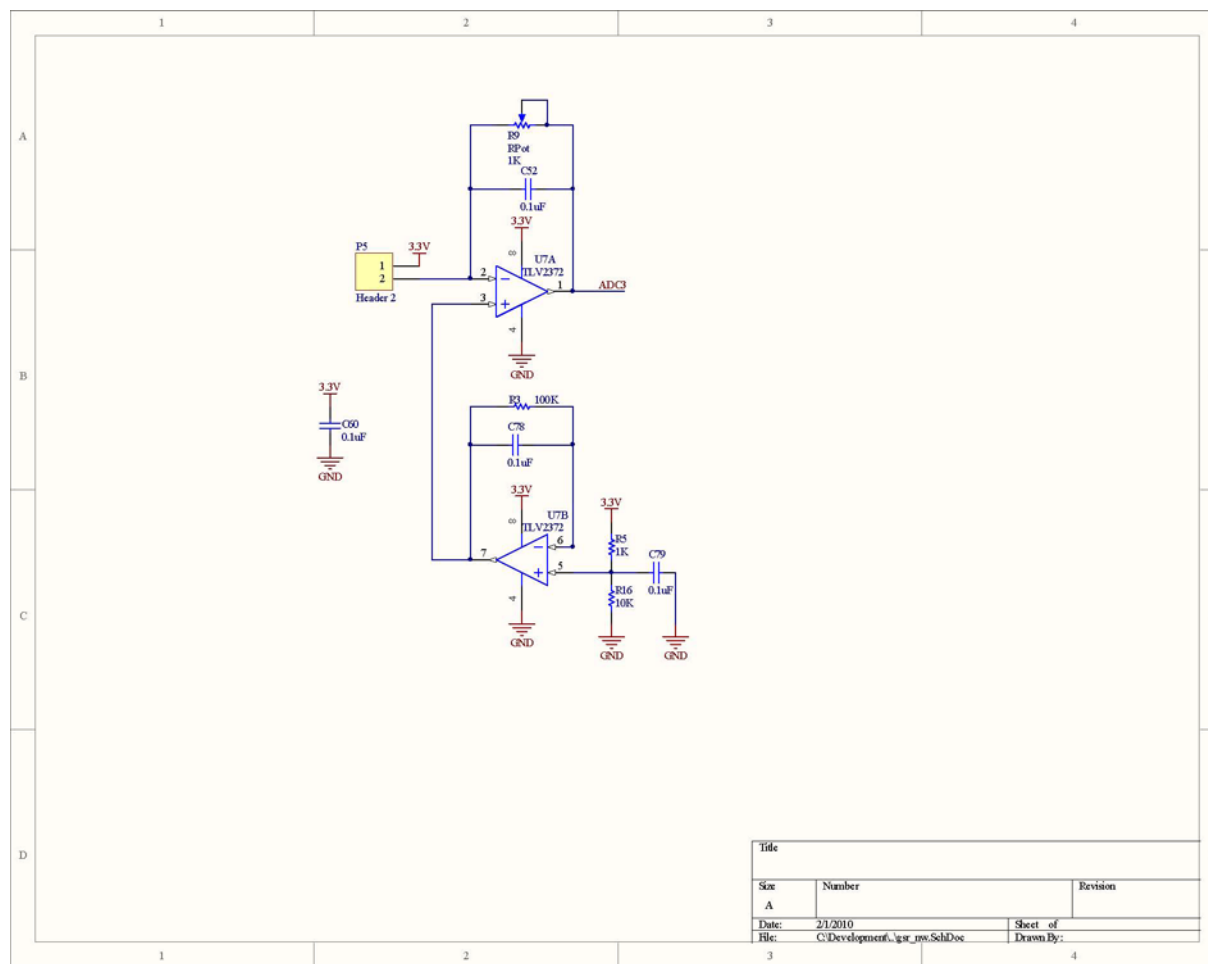
## A.1.11 Power Supply



### A.1.12 Green Board

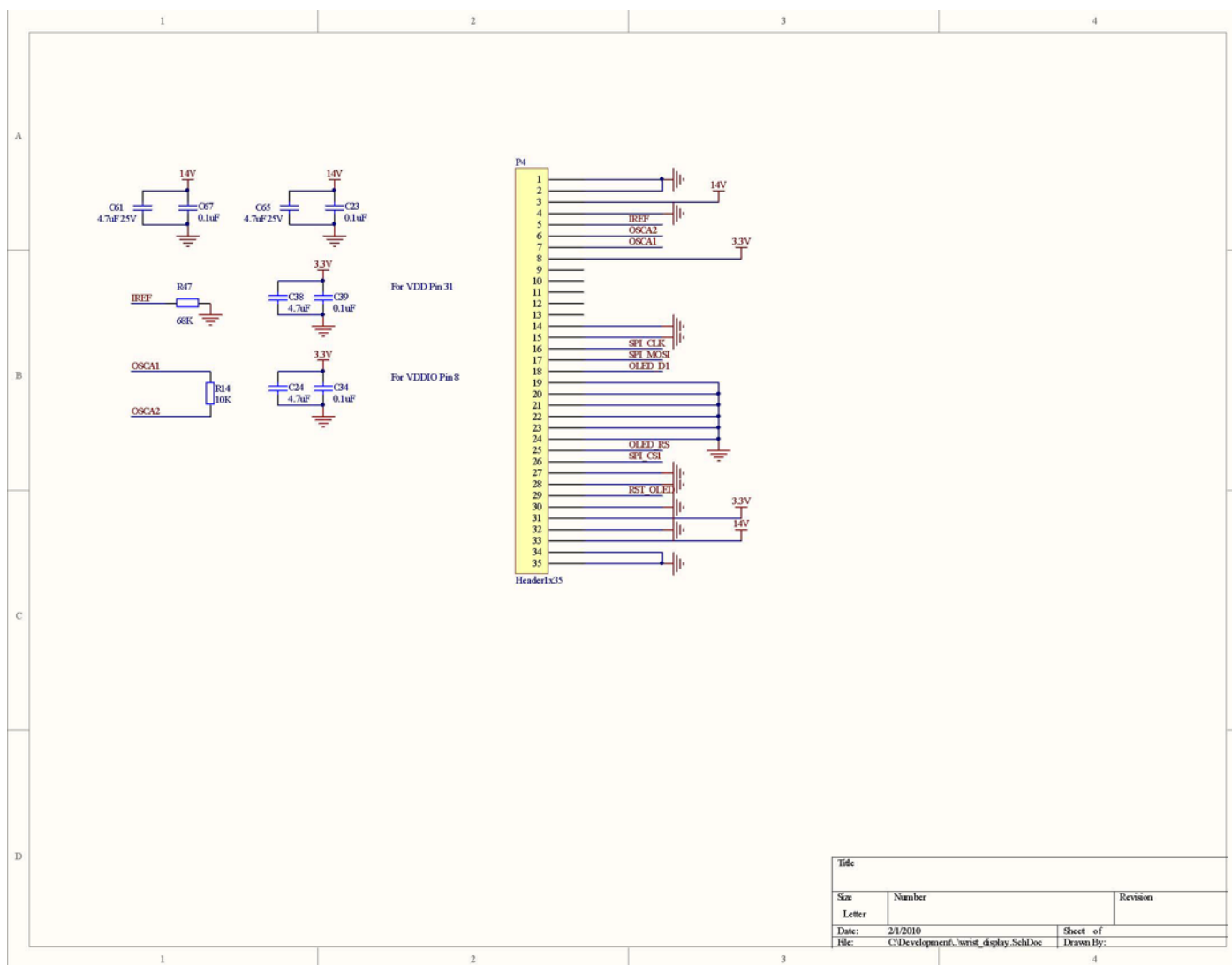
Schematics are owned by Empower Technologies.

### A.1.13 Wrist - Galvanic Skin Response Sensor

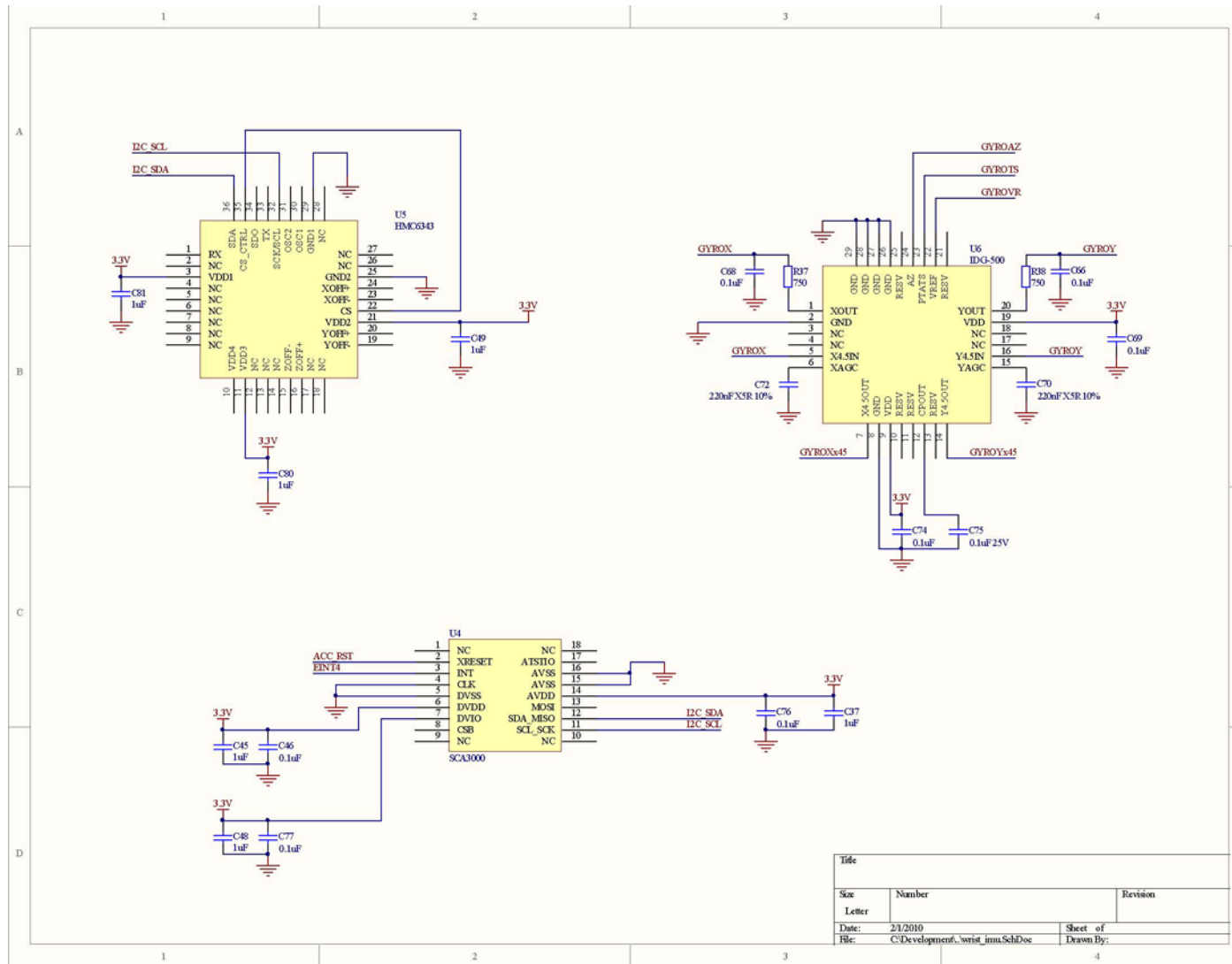




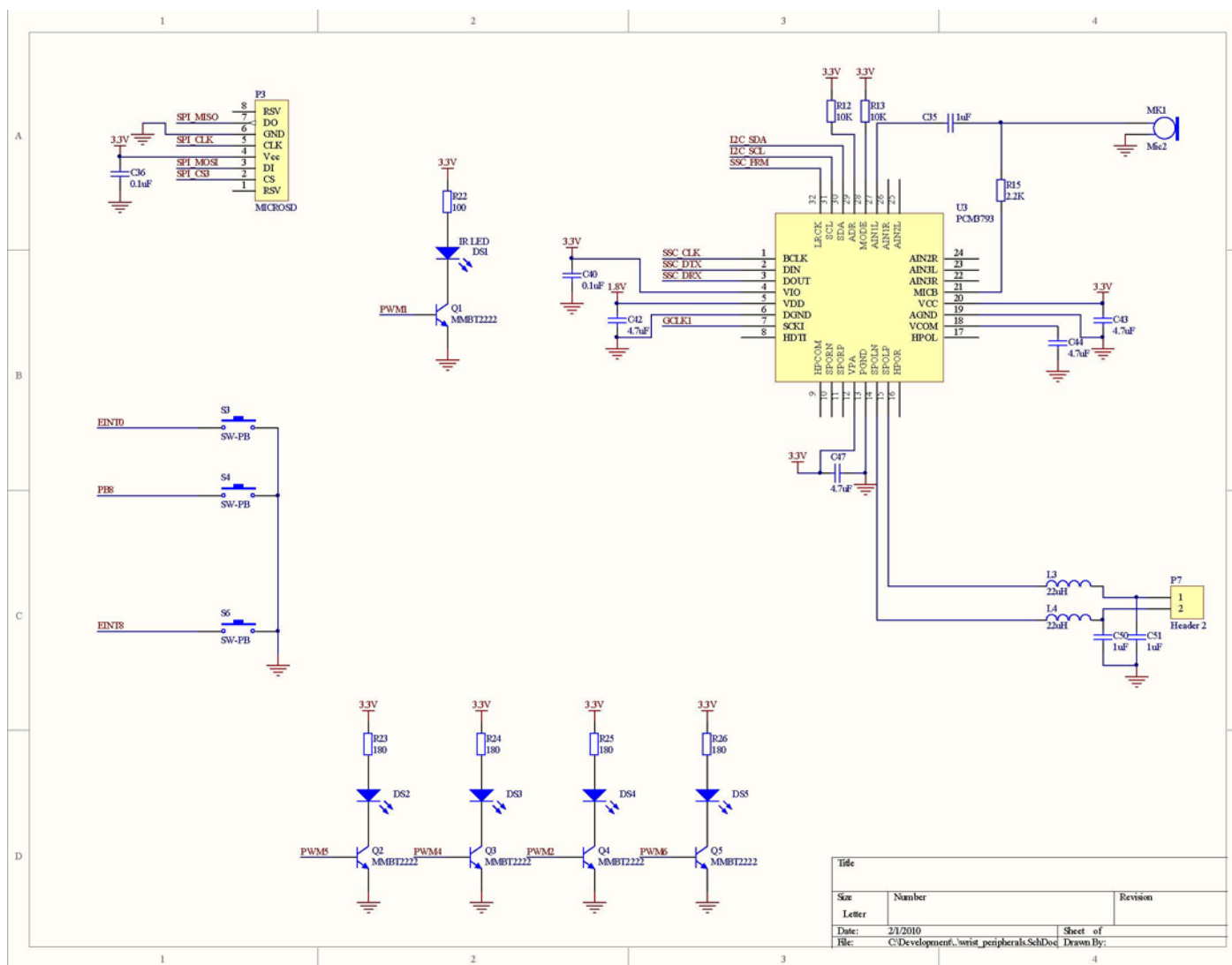
## A.1.15 Wrist - OLED Display



## A.1.16 Wrist - Inertial Measurement Unit

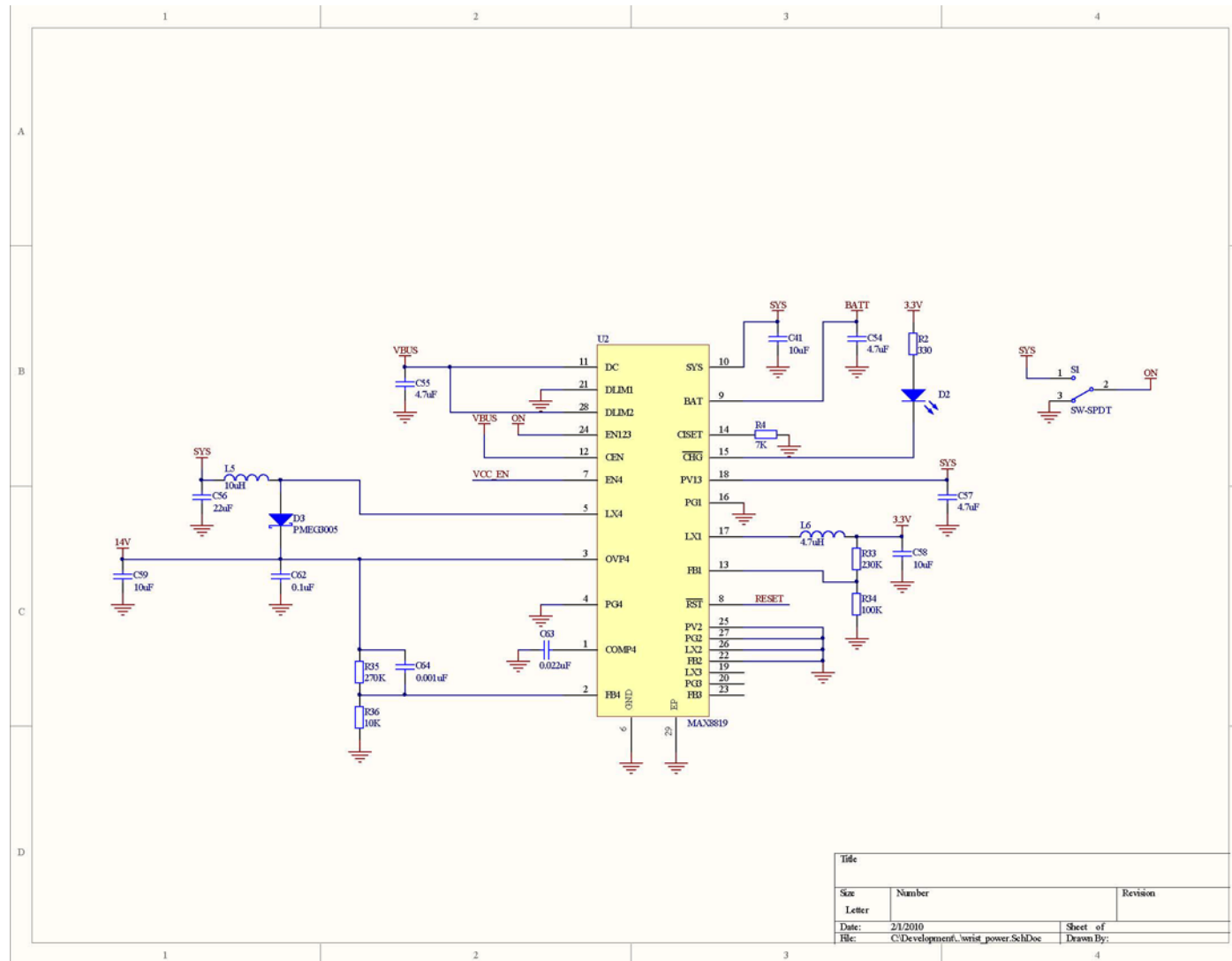


## A.1.17 Wrist - Peripherals

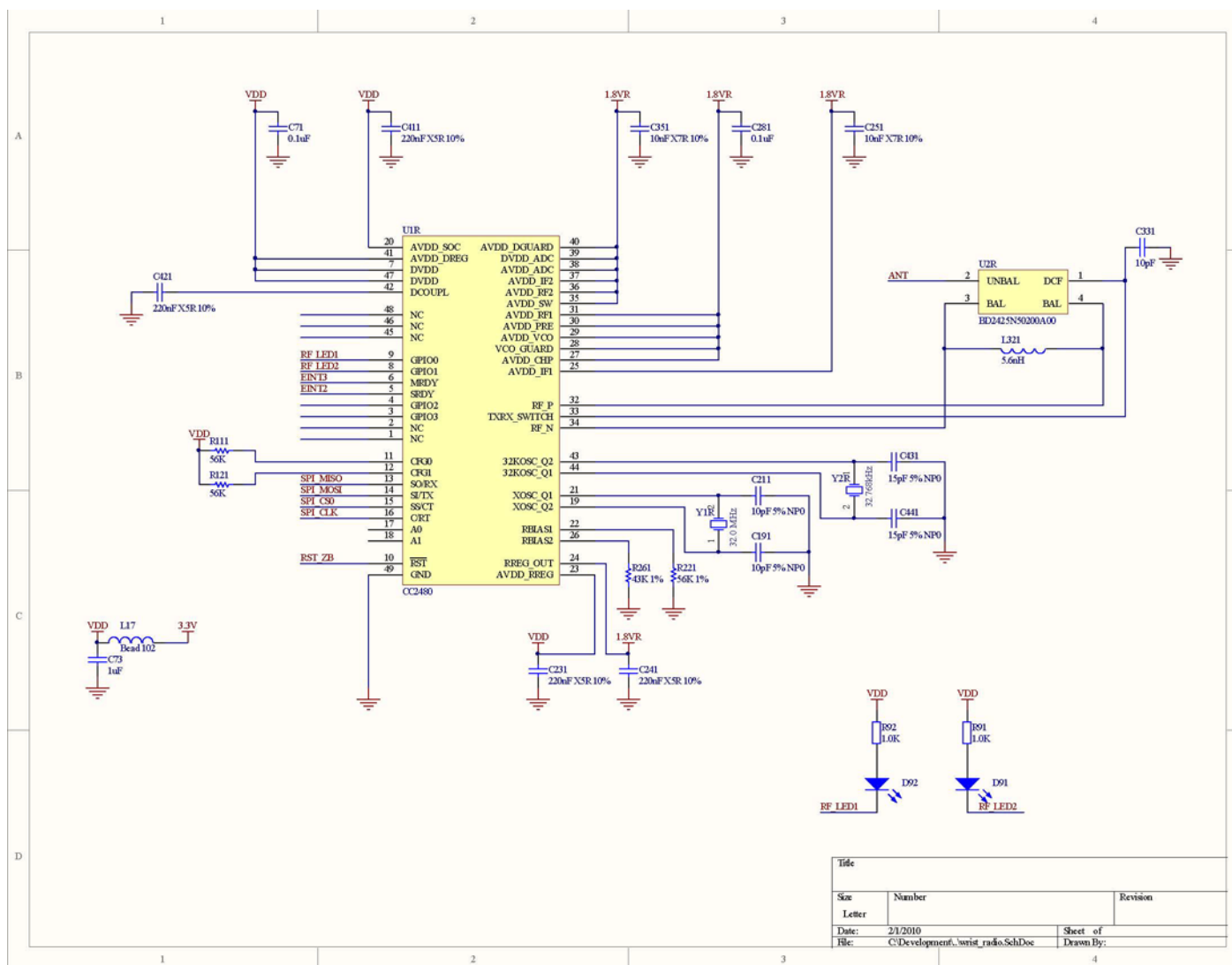




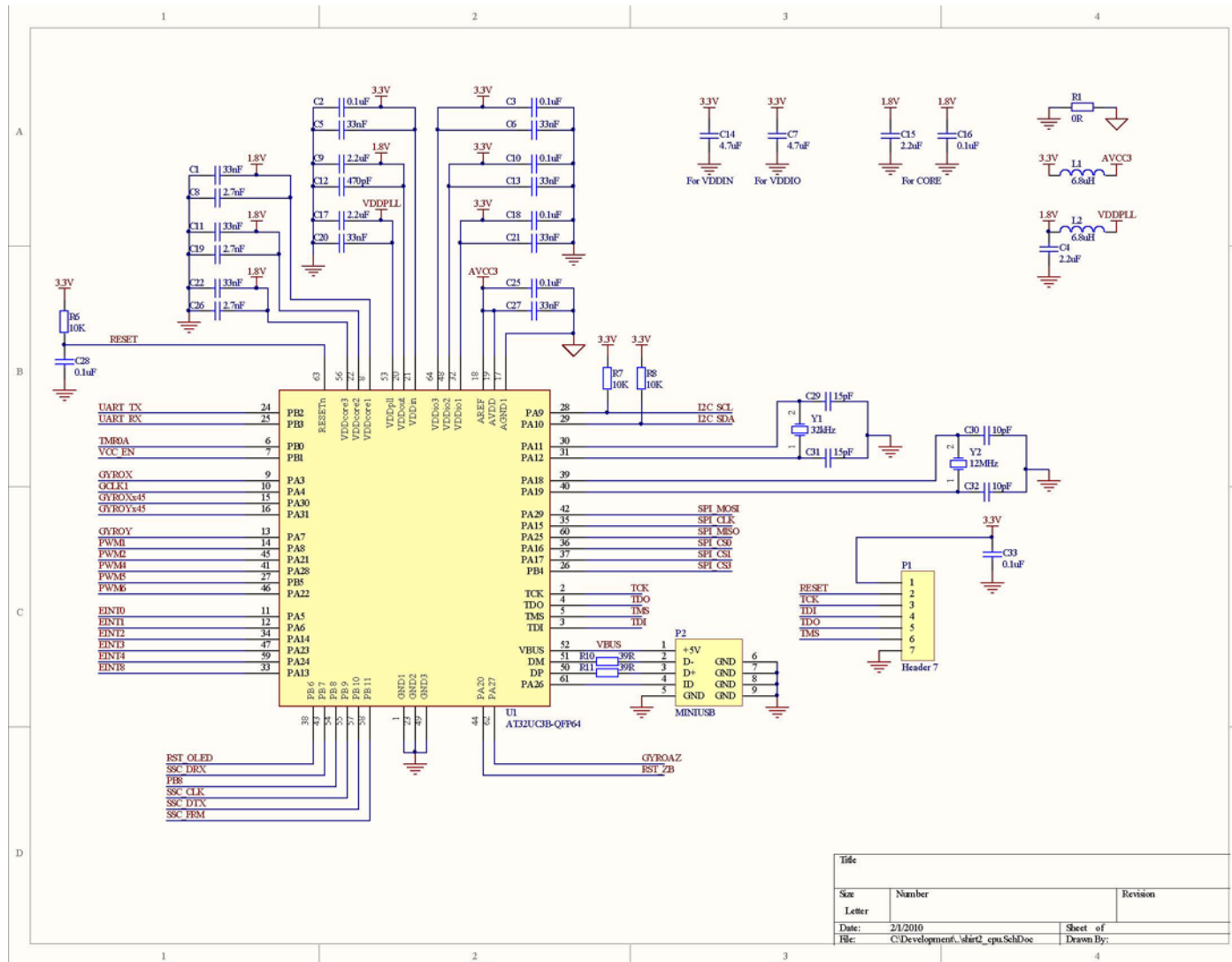
### A.1.18 Wrist - Power Supply, Battery Charge Controller



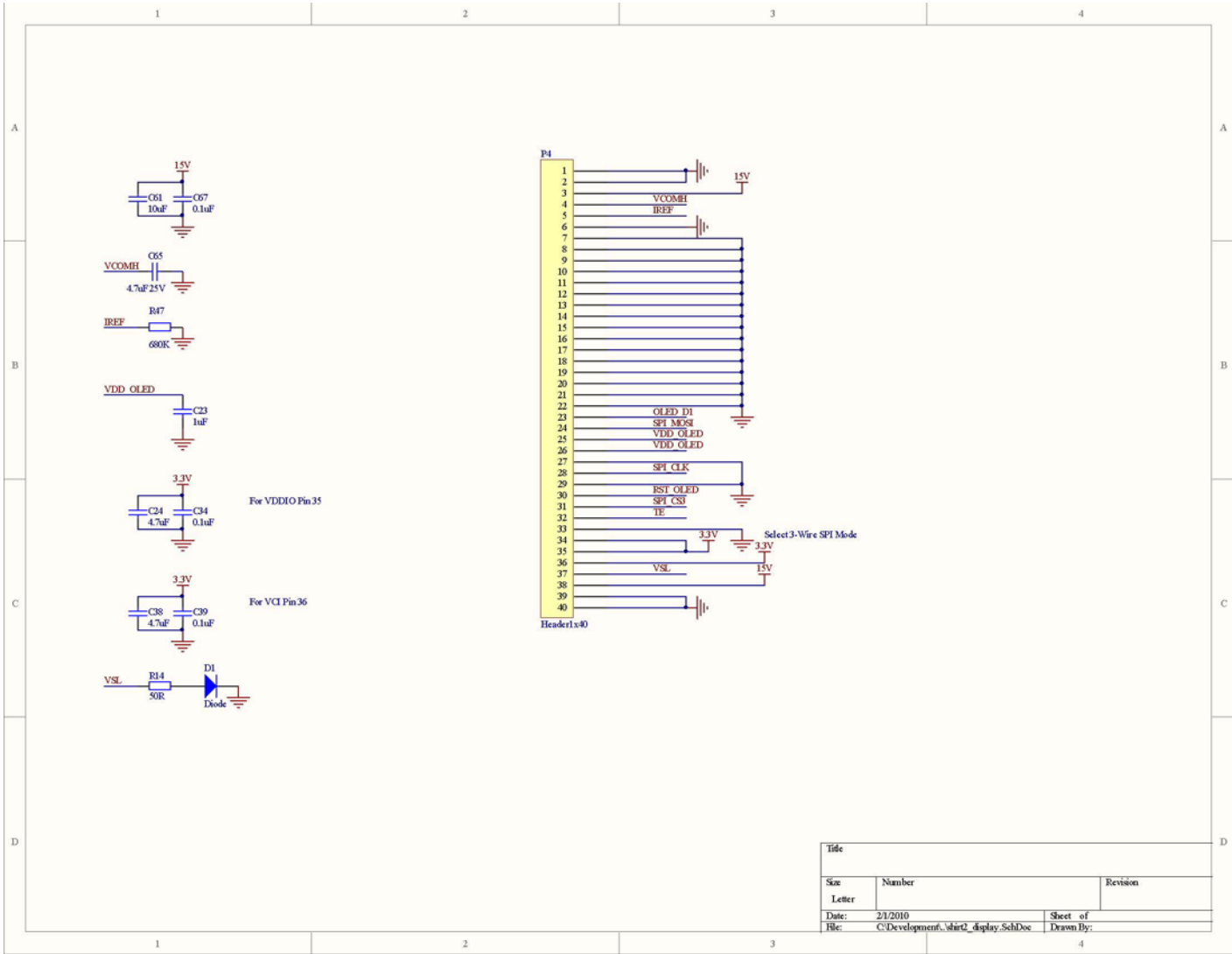
## A.1.19 Wrist - Radio



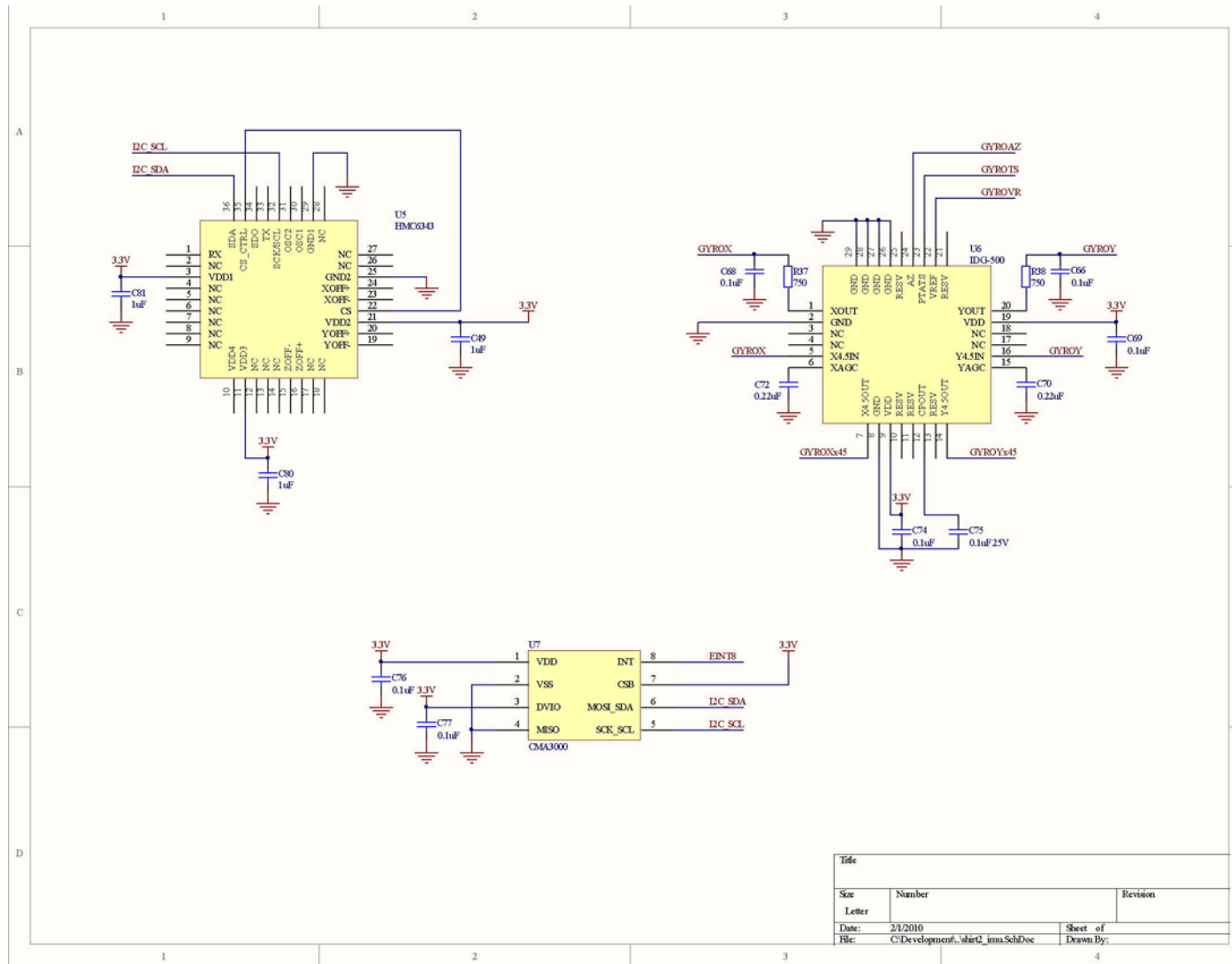
### A.1.20 μBadge - Processor



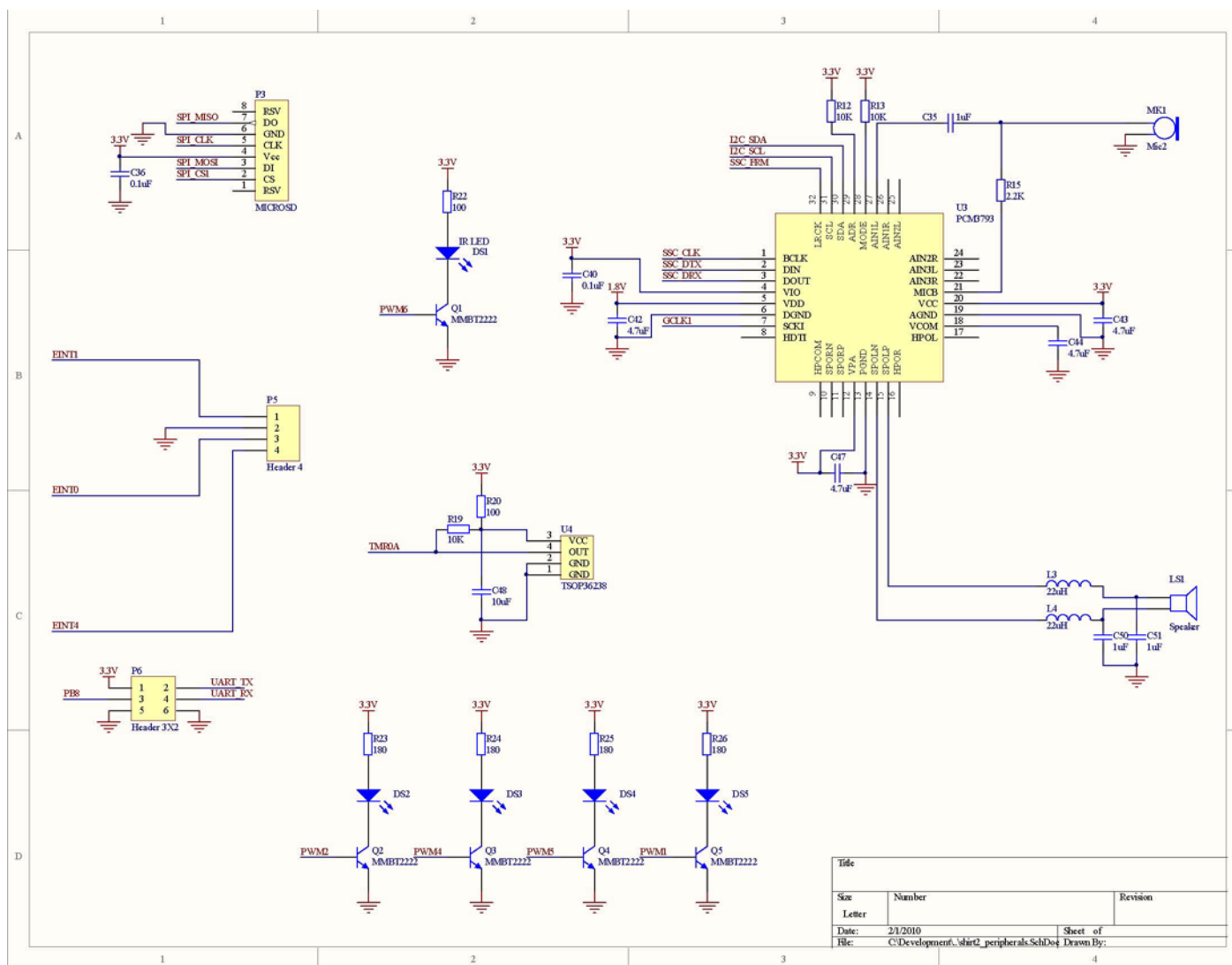
### A.1.21 μBadge - OLED Display



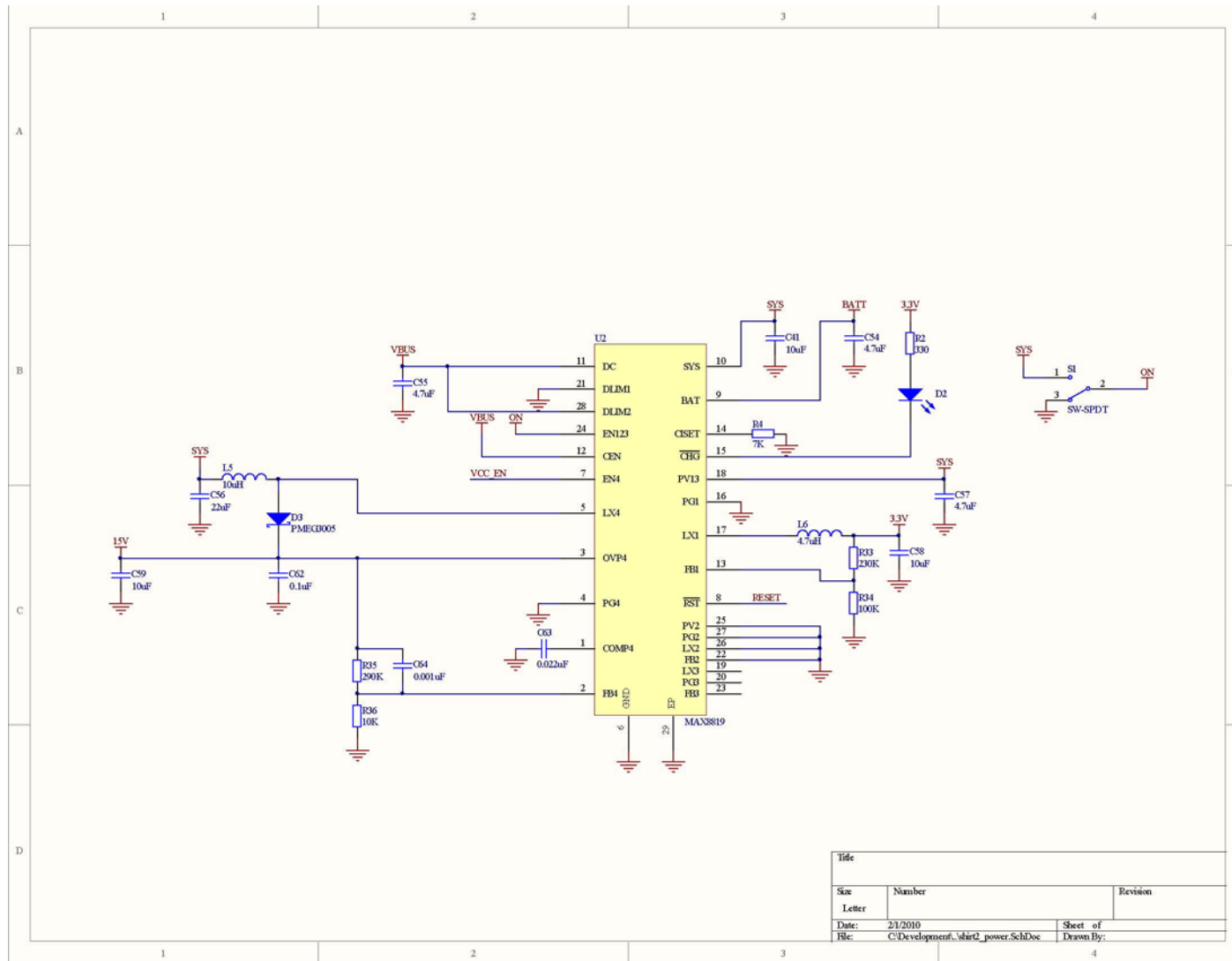
## A.1.22 µBadge - Inertial Measurement Unit



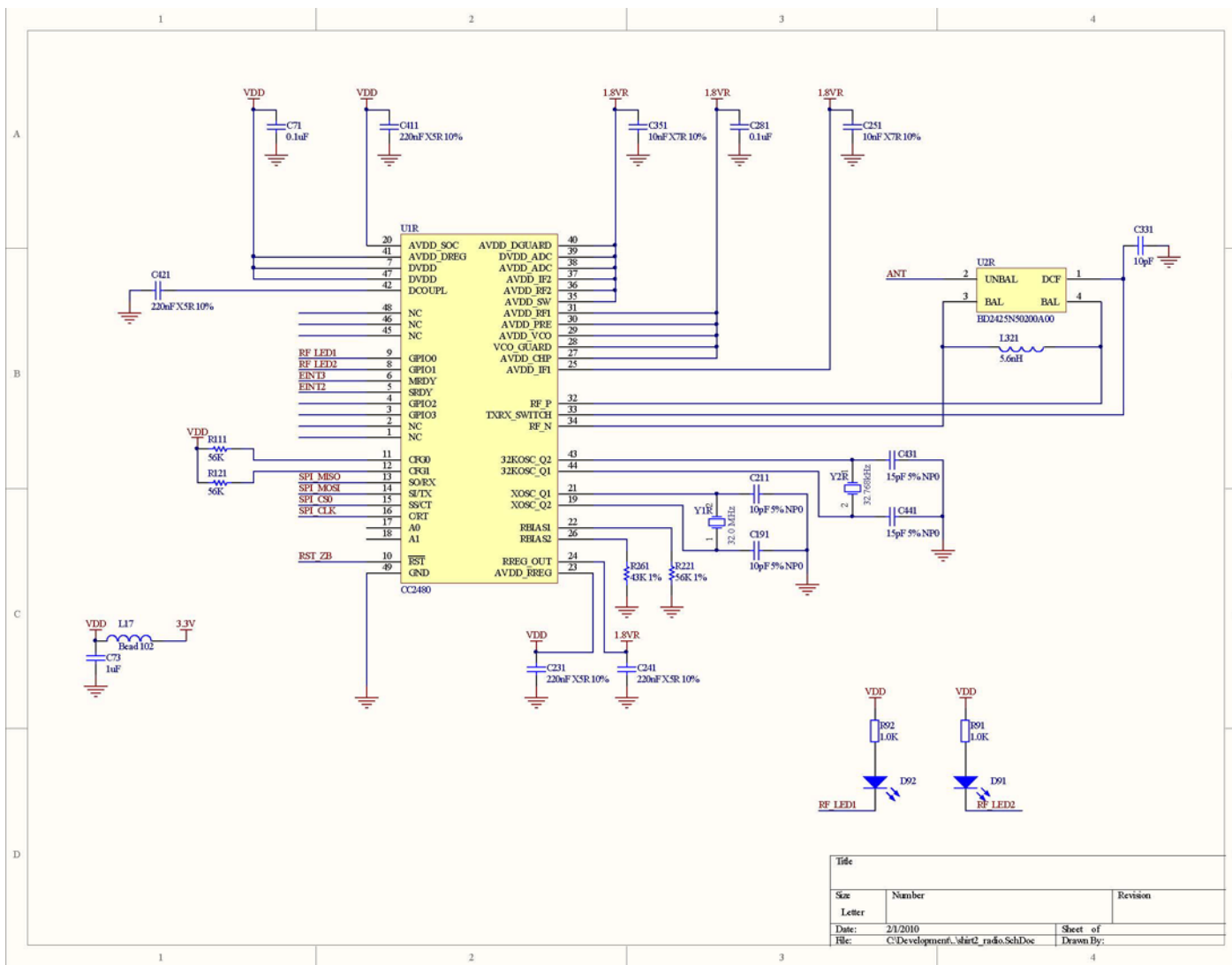
### A.1.23 $\mu$ Badge - Peripherals



## A.1.24 $\mu$ Badge - Power Supply, Battery Charge Controller



### A.1.25 $\mu$ Badge - Radio





## A.2 SOURCE CODE

---

Due to the enormous amount of code developed through the course of this dissertation, all of the source code and development tools are available at the SPINNER webpage: <http://www.media.mit.edu/resenv/spinner/>

# REFERENCES

.....

.....

I t's the same old 'songs of pain' ... Gone public domain.

- Daniel Johnston [75]

The *References* chapter provides the bibliographic information cross-referenced to the text.

- [1] Coil - Scatology (1984) Vinyl. Force And Form, FFK1. UK.
- [2] Arendt, H., *Men in dark times*. [1st ed. 1968, New York,: Harcourt. x, p. 147
- [3] Borges, J.L., *El Aleph*. 1949, Buenos Aires,: Editorial Losada. 146 p.
- [4] Vinge, V., *Rainbows end*. 1st ed. 2006, New York: Tor. 364 p.
- [5] Greenfield, A., *Everyware : the dawning age of ubiquitous computing*. 2006, Berkeley, CA: New Riders. iii, 267 p.
- [6] Weiser, M., *The computer for the 21st century*. SIGMOBILE Mob. Comput. Commun. Rev., 1999. 3(3): p. 3-11.
- [7] Feij, C., et al., *Exploring a heterogeneous and fragmented digital ecosystem: Mobile content*. Telemat. Inf., 2009. 26(3): p. 282-292.
- [8] Napoli, P.M., *Audience economics : media institutions and the audience marketplace*. 2003, New York: Columbia University Press. x, 235 p.
- [9] Bill Joy. *Why the future doesn't need us*. Wired, April 2000.
- [10] Wimmer, R.D., *Mass Media Research : an introduction*. 9th Ed. ed. 2009, Boston, MA: Cengage- Wadsworth.
- [11] Keen, A., *The cult of the amateur : how today's internet is killing our culture*. 1st ed. 2007, New York: Doubleday/Currency. 228 p.
- [12] Grossman, L. 'Time's Person of the Year: You'. Time Magazine, 13 Dec. 2006.
- [13] Handel, Jonathan, *Uneasy Lies the Head that Wears the Crown: Why Content's Kingdom is Slipping Away* (May 1, 2009). Vanderbilt Journal of Entertainment and Technology Law, Vol. 11, No. 3, pp. 597-636, 2009.

- [14] O'Reilly, Tim, *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Communications & Strategies, No. 1, p. 17, First Quarter 2007.
- [15] Larry Dobrow, *Is It Time to Put an End to Brand Integration?* ADVERTISING AGE (21 May 2009)
- [16] Baeza-Yates, R., *User generated content: how good is it?*, in Proceedings of the 3rd workshop on Information credibility on the web. 2009, ACM: Madrid, Spain. p. 1-2.
- [17] Cheliotis, G., *From open source to open content: Organization, licensing and decision processes in open cultural production*. Decis. Support Syst., 2009. 47(3): p. 229-244.
- [18] McLuhan, M. and Q. Fiore, *The medium is the message: An Inventory of Effects*. 1967, New York,: Random House. 157 p. (chiefly illus.).
- [19] McLuhan, M. and W. Watson, *From cliché to archetype*. 1970, New York,: Viking Press. 213 p.
- [20] Borst, I., Van Den Ende, J. (2008). *Motivation and contributions in crowdsourced activities*. Conference proceedings, International Product Development Management Conference. Hamburg, Germany, 30 June – 2 July 2008.
- [21] Deci, E.L. and R.M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Perspectives in social psychology. 1985, New York: Plenum. xv, 371 p.
- [22] Vaknin, S. and L. Rangelovska, *Malignant self love : narcissism revisited*. 1st ed. 2005, Prague: Narcissus Publications. 562 p.
- [23] Annenberg School of Communications (University of Southern California). Norman Lear Center., *Artists, technology & the ownership of creative content : creative control in the digital age : scenarios for the future*. 1st ed. Creativity commerce & culture. 2003, Los Angeles, CA: Norman Lear Center at the USC Annenberg School for Communication. 144 p.

- [24] Orwell, G., *Nineteen eighty-four*, a novel. 1st American ed. 1949, New York,: Harcourt. 314 p.
- [25] Daniel Soar, *Short Cuts*. LRB volume 30 number 16, 14 August 2008.
- [26] Brin, David (December, 1996). "The Transparent Society". Wired 4.12
- [27] Mathew Laibowitz, Joseph A. Paradiso: *Parasitic Mobility for Pervasive Sensor Networks*. Pervasive 2005: 255-278
- [28] N.-W. Gong, "Configurable Dynamic Privacy for Pervasive Sensor Networks," master's thesis, MIT Media Lab, 2009.
- [29] Ricœur, P., *Time and narrative*. 1984, Chicago: University of Chicago Press.
- [30] R Kearney. *On Stories*. Routledge, 2002.
- [31] Philip Page. *Furrowing all the brows: Interpretation and the transcendent in toni morrison's paradise*. African American Review, 35(4):637–649, 2001.
- [32] Vladimir A. Propp. *Morphology of the Folktale*. Publications of the American Folklore Society. University of Texas Press, 1928.
- [33] Snyder, B., *Save the cat! : the last book on screenwriting you'll ever need*. 2005, Studio City, CA: M. Wiese Productions. xvi, 195 p.
- [34] Freytag, G., *Technique of the drama; an exposition of dramatic composition and art*. 1968, New York,: B. Blom. ix, 395 p.
- [35] Vogler, C., *The writer's journey : mythic structures for storytellers and screenwriters*. 1992, Studio City, CA: M. Wiese Productions. viii, 289 p.
- [36] Ballard, J.G., *Crash*. 1973, London,: Cape. 224 p.
- [37] Chee-Yee Chong and S.P. Kumar. *Sensor networks: evolution, opportunities, and challenges*. Proceedings of the IEEE, 91(8):1247–1256, Aug. 2003.

- [38] B. Warneke, M. Last, B. Liebowitz, and K.S.J. Pister. *Smart dust: communicating with a cubic-millimeter computer*. Computer, 34(1):44–51, Jan. 2001.
- [39] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. *Passive capture and ensuing issues for a personal lifetime store*. In CARPE'04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences, pages 48–55, New York, NY, USA, 2004. ACM Press.
- [40] <http://www.vicon.com/SenseCam>
- [41] Burak Ozer Wayne Wolf and Tiehan Lv. *Smart cameras for embedded systems*. In IEEE Computer, volume 35, pages 48–53, September 2002
- [42] T. Teixeira, D. Lymberopoulos, E. Culurciello, Y. Aloimonos and A. Savvides, *A Lightweight Camera Sensor Network Operating on Symbolic Information*, Proceedings of First Workshop on Distributed Smart Cameras 2006, held in conjunction with ACM SenSys 2006
- [43] D. Lymberopoulos, T. Teixeira and A. Savvides, *Macroscopic Human Behavior Interpretations Using Distributed Imagers and Other Sensors*, under review for publication in Transactions of IEEE
- [44] T. Ko, Z. Charbiwala, S. Ahmadian, M. Rahimi, M. Srivastava, S. Soatto, and D. Estrin “*Exploring Tradeoffs in Accuracy, Energy and Latency of Scale Invariant Feature Transform in Wireless Camera Networks*” First ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2007), Vienna, Austria, September 25-28, 2007.
- [45] Mohammad Rahimi, Shaun Ahmadian, David Zats, Juan Garcia, Mani Srivastava, Deborah Estrin, “*Magic of Numbers in Networks of Wireless Image Sensors*,” Workshop on Distributed Smart Cameras (DSC), Boulder Colorado, 2006.
- [46] Snibbe, S. S. and Raffle, H. S. 2009. *Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits*. In Proceedings of the 27th international Conference on Human Factors in

Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1447-1456.

- [47] Bellotti, V. and Sellen, A.: *Design for privacy in ubiquitous computing environments*. In Proceedings of the 3rd European Conference on Computer-Supported Cooperative Work (ECSCW'93). Kluwer Academic Publishers, Norwell, MA, pp.77-92 (1993)
- [48] Wu chi Feng, Brian Code, Ed Kaiser, Mike Shea, and Wu chang Feng. *Panoptes: scalable low-power video sensor networking technologies*. In MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, pages 90–91, New York, NY, USA, 2003. ACM Press
- [49] Bell, G. and Gemmell, J., “A Digital Life,” Scientific American, 296(3), February 2007, pp. 58-65.
- [50] T. Choudhury and A. Pentland. *Sensing and modeling human networks using the sociometer*. In Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on, pages 216–222, 18-21 Oct. 2005.
- [51] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.
- [52] M. Laibowitz, J. Gips, R. Aylward, A. Pentland, and J.A. Paradiso. *A sensor network for social dynamics*. In Information Processing in Sensor Networks, 2006. IPSN 2006. The Fifth International Conference on, pages 483–491, 19-21 April 2006.
- [53] Olguín Olguín, D., et al, “Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior.” IEEE Trans. on Systems, Man, and Cybernetics - Part B, 39(1), Feb. 2009, pp. 43-55.
- [54] Push Singh and William Williams. *Lifenet: a propositional model of ordinary human activity*. In Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at K-CAP 2003, 2003.

- [55] Daniel S. Weld, Corin Anderson, Pedro Domingos, Oren Etzioni, Krzysztof Gajos, Tessa Lau, and Steve Wolfman. *Automatically personalizing user interfaces*. In IJCAI03, Acapulco, Mexico, August 2003.
- [56] Guensler R. Wolf J and Bachman W. *Elimination of the travel diary: an experiment to derive trip purpose from gps travel data*. In Transportation Research Board 80th annual meeting, 2001.
- [57] B. Logan, J. Healey, Matthai Philipose, E. Munguia Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in Proceedings of the International Conference on Ubiquitous Computing, vol. LNCS 4717. Berlin Heidelberg: Springer-Verlag, 2007, pp. 483–500.
- [58] David Bannach, Oliver Amft, Paul Lukowicz, "Rapid Prototyping of Activity Recognition Applications," IEEE Pervasive Computing, pp. 22-31, April-June, 2008
- [59] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification* (2nd Edition). Wiley-Interscience, November 2000.
- [60] E. Wertheim. *Historical background of organizational behavior*.
- [61] V. M. Bove and J. Mallett. *Collaborative knowledge building by smart sensors*. BT Technology Journal, 22(4):45–51, 2004.
- [62] J. Mallett and Jr. Bove, M.V. *Eye society*. In Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on, volume 2, pages II–17–20vol.2, 6-9 July 2003. 26
- [63] Bo Morgan and Push Singh. *Elaborating sensor data using temporal and spatial commonsense reasoning*. In BSN '06: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06), pages 187–190, Washington, DC, USA, 2006. IEEE Computer Society.
- [64] Rakoff, D., *Fraud : essays*. 1st ed. 2001, New York: Doubleday. 228 p.
- [65] June Noble. *Steal This Plot: A Writer's Guide to Story Structure and Plagiarism*. Paul S. Eriksson, 1985.



- [66] J. Meehan. *Tale-spin*. In R. Schank and C. Reisbeck, editors, *Inside Computer Understanding: Five Programs Plus Miniatures*, pages 197–226. Lawrence Erlbaum, Englewood Cliffs, NJ, 1981.
- [67] Sheldon Klein, J. F. Aeschlimann, D. F. Balsiger, S. L. Converse, C. Court, M. Foster, R. Lao, J. D. Oakely, and J. D. Smith. *Automatic novel writing*. In W. Burghardt and K. Hölker, editors, *Text Processing/Textverarbeitung*, pages 338–412. Walter de Gruyter, Berlin and New York, 1979. (Published version of UWCS Technical Report No. 186, 1973).
- [68] Ali Mazalek and Glorianna Davenport. *A tangible platform for documenting experiences and sharing multimedia stories*. In ETP '03: Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence, pages 105–109, New York, NY, USA, 2003. ACM Press.
- [69] Aisling Kelliher. *Everyday cinema*. In SRMC '04: Proceedings of the 1st ACM workshop on Story representation, mechanism and context, pages 59–62, New York, NY, USA, 2004. ACM Press.
- [70] Hyun-Yeul Lee. *Provenance as narrative history*. MIT Media Lab General Exams.
- [71] Roger C. Schank and Robert P. Abelson. *Scripts, plans, goals, and understanding*. Lawrence Erlbaum, 1977.
- [72] W Lehnert. *Plot units: A narrative summarization strategy*. Strategies for natural language processing, pages 375–414, 1982.
- [73] Andrew Steven Gordon. *The design of knowledge-rich browsing interfaces for retrieval in digital libraries*. PhD thesis, 1998. Adviser-Roger C. Schank.
- [74] Ryan Williams, Barbara Barry, and Push Singh. *Comickit: acquiring story scripts using common sense feedback*. In IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces, pages 302–304, New York, NY, USA, 2005. ACM Press.
- [75] Johnston, D. *Lost in My Infinite Memory*. Is and Always Was. 2009, Eternal Yip Eye Music, United States.

- [76] Cronenberg, D. and C. Rodley, *Cronenberg on Cronenberg*. Rev. ed. 1993, London Boston: Faber and Faber. xxiv, 214 p.
- [77] <http://focus.ti.com/docs/prod/folders/print/tms320dm6446.html>
- [78] Sayed Ahmad; Rasit Eskicioglu; Peter Graham, "Design and Implementation of a Sensor Network Based Location Determination Service for use in Home Networks," Mobile Adhoc and Sensor Systems (MASS), 2006 IEEE International Conference on , vol., no., pp.622-626, Oct. 2006
- [79] Dick Tracy from Colorado to Nova Scotia (1933, reprints May 22, 1933 to August 2, 1933 newspaper strip). Chester Gould.
- [80] Hong, J. I. and J. A. Landay (2004). *An architecture for privacy-sensitive ubiquitous computing*. Proceedings of the 2nd international conference on Mobile systems, applications, and services. Boston, MA, USA, ACM: 177-189.
- [81] Iachello, G., K. N. Truong, et al. (2006). *Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world*. Proceedings of the SIGCHI conference on Human Factors in computing systems. Montreal, Quebec, Canada, ACM: 1009-1018.
- [82] Lifton, J., M. Laibowitz, Drew Harry et al. (2009). "*Metaphor and Manifestation Cross-Reality with Ubiquitous Sensor/Actuator Networks*." IEEE Pervasive Computing 8(3): 24-33.
- [83] M. Mittal, "*Ubicorder: A Mobile Interface to Sensor Networks*," master's thesis, MIT Media Lab, 2008.
- [84] Lifton, J. and Paradiso, J.A., *Dual Reality: Merging the Real and Virtual*. in Proceedings of the First International ICST Conference on Facets of Virtual Environments (FaVE), Berlin, Germany, 27-29 July 2009
- [85] [http://www.youtube.com/watch?v=4QkBK\\_usNp8](http://www.youtube.com/watch?v=4QkBK_usNp8)
- [86] <http://www.blender.org>

- [87] Oksana Bulgakowa dir., *The Factory of Gestures / Body Language in Film*. PPMedia and Stanford Humanities Lab. DVD. 160 minutes. 2007.
- [88] Mark Knapp (1980). *Essentials of nonverbal communication*. Harcourt College Publishers.
- [89] Van den Stock, J., & de Gelder, B. (2006). *Body language influences perception of facial expression and voice prosody*. Journal of Vision, 6(6):1064.
- [90] Pentland, A. (2008). *Honest Signals: How They Shape Our World*, The MIT Press.
- [91] Argyle, M. (1975). *Bodily communication*. London, Methuen.
- [92] Alex Pentland, "A Computational Model of Social Signalin," Pattern Recognition, International Conference on, pp. 1080-1083, 18th International Conference on Pattern Recognition (ICPR'06) Volume 1, 2006
- [93] Demaris, A. and White, V. (2004) *First Impressions, what you don.t know about how others see you*. Bantom Books
- [94] Laibowitz, M., and Paradiso, J.A., "*The UbER-Badge, A Versatile Platform at the Juncture Between Wearable and Social Computing*," in Fersha, A., Hortner, H., Kostis, G. (eds), *Advances in Pervasive Computing*, Oesterreichische Computer Gesellschaft, 2004, pp.363--368.
- [95] McKee, R. 1997. *Story: Substance, Structure, Style, and the Principles of Screenwriting*. New York, NY: HarperCollins.
- [96] Snyder, B., *Save the cat! Goes to the Movies*. 2007, Studio City, CA: M. Wiese Productions. xvi, 287 p.
- [97] Lévi-Strauss, C. (1979). *Myth and meaning*. New York: Schocken Books.
- [98] Baroni, R. (2007) *La Tension narrative. Suspense, curiosité, surprise*, Paris, Seuil, coll. « Poétique ». Foreword by Jean-Marie Schaeffer.

- [99] *Vonnegut Unbound: The master of irreverence on life, death, God, humanism, and the souls of aspiring artists*, By Christopher R. Blazejewski, The Harvard Crimson, Friday, May 12, 2000
- [100] <http://sivers.org/drama>
- [101] Winerman, Lea. ' *Thin slices*' of life. Monitor Staff. March 2005, Vol 36, No. 3. Print version: p 54.
- [102] Antonioni, Michelangelo. *Michelangelo Antonioni: An Introduction*. Trans. by Scott Sullivan. New York: Simon and Schuster, 1963.
- [103] Picard, R.W., Daily, S.B. "Evaluating Affective Interactions: Alternatives to Asking What Users Feel," CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches, April 2005, Portland, OR.
- [104] *The Complete Works of Aristotle*: The Revised Oxford Translation. Ed. Jonathan Barnes. 2 vols. Princeton, NJ: Princeton Univ. Press, 1984.
- [105] <http://www.youtube.com/watch?v=rDv7D2xeKrs>
- [106] <http://www.youtube.com/watch?v=vw9outkeR1k>
- [107] <http://www.youtube.com/watch?v=ME1a0gSvMD8>
- [108] [http://www.youtube.com/watch?v=Wa4dXm7\\_yAE](http://www.youtube.com/watch?v=Wa4dXm7_yAE)
- [109] Superchunk. Seed Toss. off the Album No Pocky For Kitty (1991)

