

Deployment of a Wireless Ultrasonic Sensor Array for Psychological Monitoring

Roland Cheng, Wendi Heinzelman, Melissa Sturge-Apple, and Zeljko Ignjatovic

University of Rochester, Rochester, NY 14627, USA

{rocheng,wheinzel}@ece.rochester.edu,

melissa_sturge-apple@urmc.rochester.edu, ignjatov@ece.rochester.edu

Abstract. The deployment of a wireless sensor network to monitor subjects' locations and relative distances during psychological experiments is discussed. As its primary function, an overhead array of ultrasound sensors automatically tracks a parent, child and stranger over a $4.45\text{ m} \times 4.23\text{ m}$ observation area. The array of ultrasonic-equipped nodes can resolve two point targets separated by at least 73 cm. Each sensor can detect the range to the nearest object with 20 cm resolution. Challenges with setting up the 6×6 ultrasonic imaging array and with establishing a Kalman tracking filter are also detailed. This *WiPsy* (Wireless sensors for Psychology research) system provides accurate, real-time quantitative metrics for psychological evaluation in lieu of traditional qualitative manual coding. Moreover, tracking subjects using ultrasound sensors is less error-prone than existing methods that track based on human coding of video. Overall, *WiPsy* strives to streamline data acquisition, processing, and analysis by providing previously unavailable assessment parameters.

Keywords: Ultrasound, sensor network, imaging array, *WiPsy*.

1 Introduction

The theory of attachment, the enduring emotional bond between infant and caregiver, was first developed by psychiatrist John Bowlby in the late 1960s. Based upon conceptualizations derived from the disciplines of ethology and psychology, attachment theory posits that the quality of the parent-child relationship, particularly during infancy and early childhood, has significant implications for children's socio-emotional development. Children who form strong, healthy attachments to their caregivers are also more likely to form high quality relationships later in life.

In Bowlby's theory, the attachment relationship serves the adaptive function of providing protection and support to the child, particularly in the face of distress and threat. Using the concept of behavioral systems, Bowlby theorized that in the attachment relationship, the mother is regarded as a "secure-base" from which the infant explores and habituates to the outside environment. In times of safety and security, the infant feels free to investigate his surroundings in attempts to understand and master his environment. Think of a child putting

a toy into his mouth or banging it on the floor. Bowlby would have viewed this as exploratory behavior. Under threatening or distressful conditions, the infant returns to the mother for soothing comfort and protection. Thus, the mother is the base from which the infant derives comfort and explores. Attachment researchers regard the child's use of distance negotiation with the mother under conditions of safety and threat as an important indicator of the quality of the mother-infant attachment bond.

In order to study attachment behavior, psychologists use the well-established Strange Situation Paradigm developed by Ainsworth. The paradigm calls for a child to play with toys in the presence of a parent, a stranger, both, or alone. By monitoring the child's distance negotiation under these different stress conditions, psychologists can infer the quality of the child's attachment to the mother [1]. For over 30 years, the distance between the child-parent and child-stranger have been assessed by trained human coders. The coding process is unsatisfactorily qualitative and tediously slow.

The system described in this paper replaces human coding with more accurate, quantitative measurements. During the strange situation experiments, an ultrasonic video of the observation room is recorded using a wireless array of ultrasound sensors. The ultrasound data is processed in real time through a tracking filter, which extracts the positions of the child, parent, and stranger and displays the child-parent and child-stranger distances. This gives psychologists the ability to assess behavioral distance negotiation instantaneously. Furthermore, given that the attachment relationship exists across multiple behavioral levels including physiological and psychological levels, this system allows researchers to simultaneously compare assessments across multiple levels of analysis.

2 System Architecture

The complete system architecture consists of two major components: the imaging subsystem and the tracking subsystem. The former is composed of an overhead array of 6×6 Tmote Sky nodes, each equipped with a set of Devantech SRF05 ultrasonic transmitter/receiver transducers. The array transmits all echo measurements via wireless communication to a base station, which assembles and displays a real-time image of the room using Matlab. This data is also fed into a tracking filter, which estimates the positions of three possible targets within the room. Once positions are determined, the child-parent and child-stranger distances are computed and displayed.

Instead of using conventional wired communication, the imaging system relays information over a wireless channel as an investigative procedure to determine whether fading channels have sufficient bandwidth for conveying real-time ultrasound images. Future work may include a means of easily deploying a movable, battery-operated version of the array into homes for *in situ* psychological monitoring.

Conventional wireless sensor networks often require a device be placed onto each tracked target [2, 3]. Such systems include Olivetti Research's Active

Badge [4], AT&T's Active Bat [5], Microsoft's RADAR [6], MIT's Cricket [7], and Ascension's MotionStar [8]. However, this imaging system must avoid using tracking tags on targets because child targets are often unwilling to wear them. Attempts to attach such tags result in the child's uncooperation or a forced tag removal midway through the experiment. To circumvent this limitation, the imaging system is set up as a large-scale ultrasonic array akin to conventional ultrasonic arrays used in fetal imaging. In essence, the ultrasonic array takes snapshots of the room without requiring targets to wear specialized tracking devices.

Other tracking systems that also do not require the wearing of tracking tags are computer vision systems [9, 10]. However, those systems suffer from larger processing requirements and possible occlusion problems [2]. This system requires comparatively less computation and avoids most occlusion situations.

2.1 Mote Development Environment

The Tmote Sky motes are programmed using the TinyOS development environment. TinyOS version 2 is chosen for access to the high resolution timers. These 32 kHz clocks enable the ultrasound sensors to provide a vertical height resolution of approximately 1 cm. However, in practice, these clocks are empirically slowed down by a factor of 20, which corresponds to a vertical resolution of about 20 cm. The justification for this reduced clock speed is simply to give the microprocessor sufficient CPU time to perform required computations (such as checking the radio for received packets) instead of being inundated with processing timer triggers. Since the smallest subjects, 18-month old children, typically have heights around 80 cm [11], this decreased resolution does not affect the system's detection ability.

2.2 Sensor Array

Attached to each ceiling-mounted mote is a downward-facing ultrasound sensor, shown in Figure 1. When signaled by the mote, the sensor emits an 8-cycle, 40 kHz ultrasound pulse. This directed pulse returns to the detector after being reflected off either the floor or an interposed object. The flight time of the pulse gives the range to the closest object below the sensor. Usually, this object is a subject's head, a chair, a toy on the floor, or the floor itself.

By arranging a 6×6 grid of these ultrasound sensors (shown in Figure 2), an image of the $4.45 \text{ m} \times 4.23 \text{ m}$ observation room can be taken. Locations of targets of interest are extracted from this image using peak detection. The sensor array is capable of discerning two adults separated by at least 73 cm. Objects separated by less than this minimum distance will appear to the imaging system as a single entity.

2.3 Tracking Filter

Because of the poor spatial resolution of these ultrasound images, multiple, closely clustered targets appear as a single peak instead of as separate individuals. To



Fig. 1. Ceiling-mounted mote with attached ultrasound sensor

circumvent this problem, a tracking filter is implemented to estimate the subjects' positions when the imaging system is unable to resolve all targets.

One of the simplest linear tracking filters to use is the Kalman filter. This adaptive filter minimizes the mean square error between a set of noisy measurements and some (noiseless) state space variables, assuming the state variables form a Markov chain. Some advantages with using a Kalman filter over a Wiener filter are the former's causal nature and its ability to handle non-stationary processes [12]. Therefore, the Kalman filter can be used for real-time tracking and imposes few assumptions about the data's behavior.

The vector Kalman filter implemented in this system assumes that the state space variable $\mathbf{s}[n|n]$ at current time n evolves according to a Gauss-Markov model. Specifically, the state variable of three tracked targets is represented as an 18-element tuple:

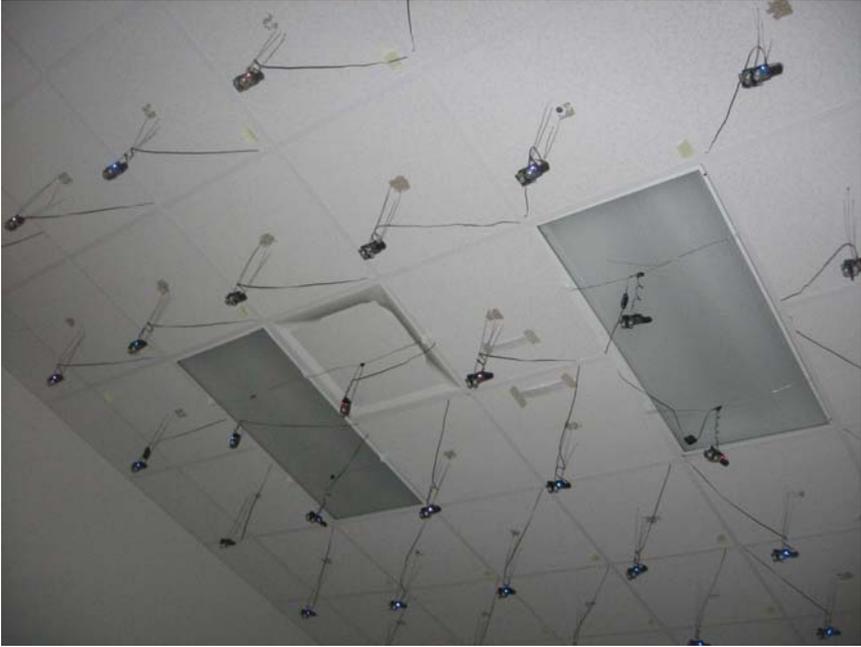
$$\mathbf{s}[n|n] = [\mathbf{s}_1[n] \ \mathbf{s}_2[n] \ \mathbf{s}_3[n]] \quad (1)$$

$$\mathbf{s}_i[n] = [x_i[n-1] \ x_i[n] \ v_{i,x}[n] \ y_i[n-1] \ y_i[n] \ v_{i,y}[n]] \quad (2)$$

where $(x_i[n], y_i[n])$ is target i 's location at time n and $v_i[n]$ is the target's velocity also at time n . The estimate $\hat{\mathbf{s}}[n|n]$ of the state variable at time n is incrementally computed using the prediction vector $\hat{\mathbf{s}}[n|n-1]$, the Kalman gain vector $\mathbf{K}[n]$, the noisy observation vector $\mathbf{w}[n]$, the minimum predicted mean square estimate matrix $\mathbf{M}[n|n-1]$, and the minimum mean square error matrix $\mathbf{M}[n|n]$. All of the previous quantities are either given or can be computed from (4), (5), (6), and (7):

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n] (\mathbf{w}[n] - \mathbf{H}\hat{\mathbf{s}}[n|n-1]) \quad (3)$$

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1] \quad (4)$$



(a) Worm's eye view



(b) Profile view

Fig. 2. 6×6 array of motes

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^T (\mathbf{C} + \mathbf{H}\mathbf{M}[n|n-1]\mathbf{H}^T)^{-1} \quad (5)$$

$$\mathbf{M}[n|n-1] = \mathbf{A}\mathbf{M}[n-1|n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \quad (6)$$

$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H})\mathbf{M}[n|n-1] \quad (7)$$

where \mathbf{A} is the state transition matrix; \mathbf{H} is the matrix transforming $\mathbf{s}[n|n]$ into $\mathbf{w}[n]$; \mathbf{C} is the covariance matrix of the white, Gaussian measurement noise added to $\mathbf{s}[n|n]$; \mathbf{B} is the matrix transforming the driving noise term $\mathbf{u}[n]$ into additive noise for $\mathbf{s}[n-1|n-1]$; \mathbf{Q} is the covariance matrix for $\mathbf{u}[n]$; and \mathbf{I} is the identity matrix. The recursion is initiated using $\hat{\mathbf{s}}[0|0] = [110110110110110110]$ and $\mathbf{M}[0|0] = \mathbf{C}$ [12]. That is, the initial state assumes that all targets are stationary and start at the door, which is located at (1,1).

In the tracking filter, \mathbf{H} and \mathbf{B} are further simplified and assumed to be identity matrices. Thus, the state variable $\mathbf{s}[n|n]$ and the noisy observation vector $\mathbf{w}[n]$ are identical in size. It is further assumed that each target's x - and y -locations operate independently and that all three targets move mutually independently. Given these independence assumptions, the covariance matrices \mathbf{C} and \mathbf{Q} , and the state transition matrix \mathbf{A} are almost diagonal matrices:

$$\mathbf{C} = \begin{bmatrix} c & 0 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & 0 & 0 & 0 \\ 0 & 0 & c & 0 & 0 & 0 \\ 0 & 0 & 0 & c & 0 & 0 \\ 0 & 0 & 0 & 0 & c & 0 \\ 0 & 0 & 0 & 0 & 0 & c \end{bmatrix} \quad (8)$$

$$\mathbf{c} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & 2\sigma^2 \end{bmatrix} \quad (9)$$

$$\mathbf{Q} = 0.001\mathbf{I} \quad (10)$$

$$\mathbf{A} = \begin{bmatrix} a & 0 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & a \end{bmatrix} \quad (11)$$

$$\mathbf{a} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & \Delta t \\ -\frac{1}{\Delta t} & \frac{1}{\Delta t} & 0 \end{bmatrix} \quad (12)$$

where $\sigma^2 = 1$, $\Delta t = \frac{1}{\text{frame rate}} = 1$ s, and the $\mathbf{0}$ entries in \mathbf{C} and \mathbf{A} are 3×3 matrices of zero. The choice of the variance value for \mathbf{Q} was empirically determined

to balance between the smoothness of tracking target movements against the slow response time of tracking fast-moving targets. With larger variance values for (10), the position estimates abruptly jump between the locations determined by the peak detectors. On the other hand, smaller variance values yield smoother position tracking at the expense of having a slow response to targets with rapidly altering velocities.

Internally, the Kalman filter assumes that targets move in simple rectilinear motion. Therefore, the state transition matrix \mathbf{A} predicts a target's current location based on its previous position and velocity, and determines the target's new velocity from the current and previous positions. Older position measurements were not used to calculate velocity to allow for fast-changing velocities. Notably, this occurs when the target alters direction or speed, which breaks the rectilinear motion assumption.

3 Surmounted Challenges

In order to create the current system, several obstacles have to be surmounted. First, the power constraint is avoided by deploying a network of USB cables and hubs. Second, two conflicting goals have to be reconciled: imaging noise quality versus imaging speed. Image noise is mitigated by sensing in a staggered access pattern and by using differential image calibration. Image noise can also be minimized by waiting for secondary reflections to attenuate, at the expense of decreased frame rates. On the other hand, imaging speed can be increased by transmitting multiple sensor readings simultaneously.

3.1 Power Network

Typically, sensor networks have severe power restrictions because of the limited charge carrying capacity of modern batteries. Because this imaging system monitors subjects during psychological experiments, a system whose power source must be frequently replaced is unacceptable. More so, the nature of the psychological experiments implies that each experimental trial can only be performed once. Thus, having a mote battery deplete in the middle of a running experiment must be avoided.

To circumvent this power constraint, a power network has to be deployed to give each sensor an unlimited power source. The most convenient course of action is to network together several commercial USB hubs, USB cables, and power strips. Such a power network can supply 5 V to each mote through existing USB ports, which are present on each Tmote Sky mote.

Although 5 V are being supplied by the power network, the output voltage supplied by the mote to the ultrasound sensor is only 3 V. Since each mote can be powered by two 1.5 V AA batteries, a mote can only send a maximum of 3 V through its output pins. As shown in Table 1, the sensors require a nominal power supply of 5 V. This 2 V shortfall actually does not drastically reduce the sensor performance. The ultrasound sensors still operate correctly at 3 V, but have a smaller object-detection range.

3.2 Sensor Access Pattern: Noise vs. Speed Tradeoff

The ultrasound sensor determines range based on the first reflection received after transmission. However, this primary reflection continues to bounce around the room long after its initial detection. These secondary reflections pose a problem if they are detected by other sensors. In particular, if an adjacent sensor hears the secondary reflection emitted from another sensor before it hears its own primary reflection, then the adjacent sensor would infer the incorrect range to the object underneath it. The detection of secondary reflections yields noisy ultrasound images.

One way to ameliorate this problem is by having only one single sensor emit a pulse at a time. If multiple sensors were to simultaneously emit ultrasound pulses, the probability of detecting a secondary reflection would be increased. Instead, by partitioning time access over the room, each sensor can send and receive an echo during its own time slot without interference from other sensors.

Secondary reflections can also be mitigated by increasing the effective spatial separation between consecutive transmitters or by increasing the time duration between transmitter emissions. The first method can be done with no penalty to the video frame rate. By increasing the distance between transmitters, a secondary echo must propagate farther and undergo more reflections before hitting another listening sensor. This additional traveling attenuates the secondary echoes until they are beneath the detection threshold. In practice, the physical distance between sensors need not be changed in order to increase the spatial separation. Instead, re-ordering the sensor access pattern from the raster order shown in Figure 3a to the staggered order shown in Figure 3b is equivalent to increasing the distance between ultrasound firings.

Contrary to increasing the spatial separation, increasing the temporal separation between consecutive transmitters decreases the video frame rate. By making each sensor time slot longer, secondary reflections are allowed more time to reverberate and attenuate before the next sensor fires. However, by doing so, the total time needed to acquire a complete frame is also increased, which leads to decreased video frame rates. To avoid crippling the frame rate, the time slot

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

(a) Raster ordered access

2	6	10	4	8	12
24	20	16	22	18	14
26	30	34	28	32	36
1	5	9	3	7	11
23	19	15	21	17	13
25	29	33	27	31	35

(b) Single-transmit, staggered access

1	13	8	2	14	7
5	15	11	6	16	12
17	9	4	18	10	3
2	14	7	1	13	8
6	16	12	5	15	11
18	10	3	17	9	4

(c) Multiple-transmit, staggered access

Fig. 3. Sensor access patterns

length T_0 is set to twice the minimum time needed to detect the primary reflection. That is, $T_0 = 2t_{\text{round-trip}} = 2\frac{2h_{\text{ceiling}}}{c} = 25$ ms, where the ceiling height h_{ceiling} is 2.13 m and the speed of sound in air c is 343 m/s. This yields a video frame rate of about 1 frame per second. Setting T_0 to twice the minimum time is an empirical compromise that decreases image noise without sacrificing too much frame rate.

The imaging frame rate can be increased by simultaneously sending ultrasound pulses from different sensors. To avoid detecting secondary reflections, the spatial separation technique can be applied without a speed penalty. Thus, a hybrid approach between sending simultaneous pulses and using spatial separation can be used to increase the frame rate without introducing too much noise. The array access order for this technique is shown in Figure 3c.

3.3 Differential Calibration

Another way to decrease image noise is by performing differential calibration. By taking a picture of the empty observation room and subtracting that image from every video frame, fixed pattern noise can be eliminated. Fixed pattern noise can arise from minor sensor-to-sensor or mote-to-mote variations, or simply from static background noise. In this case, background noise is the dominant source of fixed pattern noise, as shown in Figure 4. The figure shows a large peak due to echoes reflecting off the one-way observation mirror.

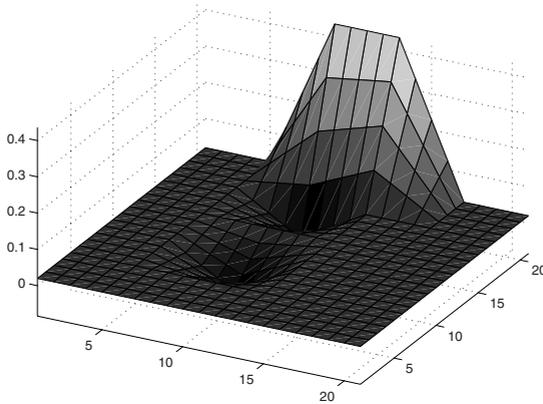


Fig. 4. Image of fixed pattern noise of the empty observation room

4 Results

A sample frame of the ultrasound images captured during the Strange Situation Paradigm is shown in Figure 5.

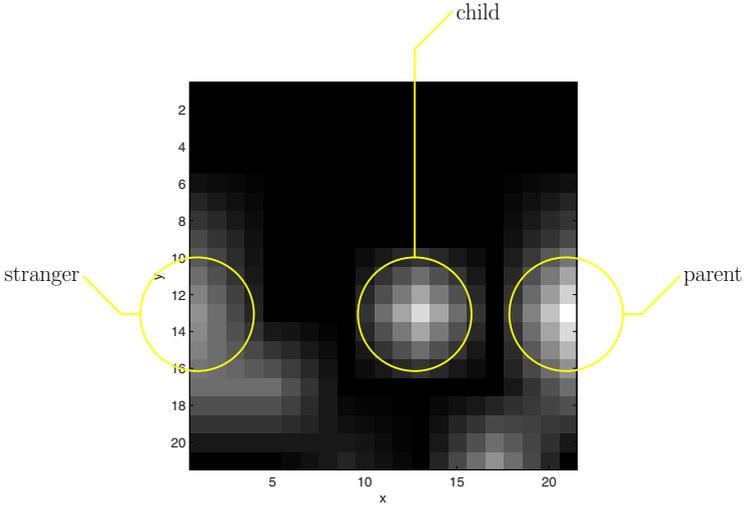


Fig. 5. Ultrasonic image of stranger, child and parent in observation room (bird's eye view)

4.1 Performance of Distance Tracking

Figure 6 compares an adult's tracked position with the sensor acquired measurements and the human-coded position readings. The mean square errors between the Kalman-filtered position estimates and the human-coded positions are $MSE_x = 0.23 \text{ m}^2$ in the x-direction and $MSE_y = 0.59 \text{ m}^2$ in the y-direction.

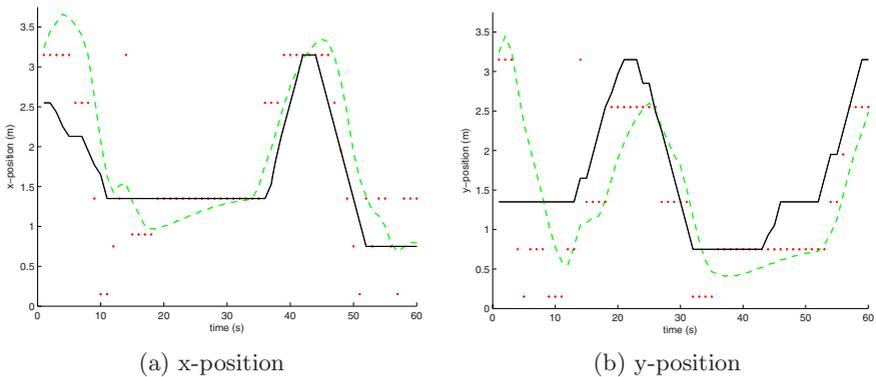


Fig. 6. XY position of tracked subject. Black solid lines are human-coded positions; dotted red points are sensor acquired positions; dashed green lines are Kalman-filtered output of the sensor measurements.

4.2 System Specifications

Relevant specifications for the sensor and the array are summarized in Table 1.

Table 1. Sensor and array specifications

	Specification	Nominal	Actual
sensor	half-power beamwidth	40°	-
	power supply	5 V	3 V
	max detectable distance	4 m	2.1 m
	resolution in propagation direction	-	20 cm
array	resolution in azimuth plane	-	73 cm × 73 cm
	sensor separation	-	61 cm × 61 cm
	observation area	-	4.45 m × 4.23 m

5 Conclusion

The WiPsy ultrasonic imaging array serves to replace human-coded distance tracking with automatic, accurate distance measurements taken during psychological experiments. A successful deployment of the sensor array requires overcoming several challenges, including creating a power network, scheduling a sensor access pattern, and using differential calibration.

6 Future Work

It is possible to extend the current work in at least two ways: utilize phase information to generate higher quality ultrasound images and track subjects using the forward-backward propagation algorithm. The current imaging system sends a series of directed ultrasound pulses to sample the scene. Instead, by using omnidirectional ultrasound pulses and processing the scattered signal at every sensor node, the signal can be amplified by an array gain. With proper signal processing, and perhaps spread-spectrum apodization, a higher-quality ultrasound image may be generated, much as is done in medical ultrasound imaging [13].

The current tracking filter is a causal Kalman filter. By removing the causality constraint, a more generalized algorithm can be used for tracking. The backward-forward propagation algorithm can use past and future information to more accurately determine the current location of a target. Such algorithms have proven quite successful in decoding digital capacity-approaching codes [14].

Acknowledgements. The lead author must acknowledge the effort of two undergraduates, Kyle Aures and Jian Chen, whose work during the summer of 2007 bootstrapped the deployment of the WiPsy ultrasonic sensor array. Furthermore, funding for this research has been provided by the National Institute of Nursing Research (R21 NR010857).

References

1. Ainsworth, M.D.S., Blehar, M.C., Waters, E., Wall, S.: *Patterns of Attachment: A Psychological Study of the Strange Situation*. Lawrence Erlbaum Associates, Hillsdale (1978)
2. Hightower, J., Borriello, G.: *Location Systems for Ubiquitous Computing*. Computer 34 (August 2001)
3. Niculescu, D., Nath, B.: *Ad Hoc Positioning System (APS)*. In: *IEEE Globe-Com*, San Antonio, TX (November 2001)
4. Want, R., Hopper, A., Falcão, V., Gibbons, J.: *The Active Badge Location System*. *ACM Transactions on Information Systems* 10 (January 1992)
5. Harter, A., Hopper, A., Steggle, P., Ward, A., Webster, P.: *The Anatomy of a Context-Aware Application*. In: *5th ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Seattle, WA (August 1999)
6. Bahl, P., Padmanabhan, V.N.: *RADAR: An In-Building RF-Based User Location and Tracking System*. In: *Proceedings of IEEE Infocom*, Tel-Aviv, Israel (March 2000)
7. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: *The Cricket Location-Support System*. In: *6th ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Boston, MA (August 2000)
8. Ascension Technology Corp., Burlington, VT, *Technical Description of DC Magnetic Trackers* (2001)
9. Atiya, S., Hager, G.: *Real-Time Vision-Based Robot Localization*. In: *Proceedings of 1991 IEEE International Conference on Robotics and Automation*, Sacramento, CA (April 1991)
10. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: *Multi-Camera Multi-Person Tracking for Easy Living*. In: *Proceedings of 3rd IEEE International Workshop on Visual Surveillance*, Dublin, Ireland (July 2000)
11. Kuczmariski, R.J., Ogden, C.L., Grummer-Strawn, L.M., Flegal, K.M., Guo, S.S., Wei, R., Mei, Z., Curtin, L.R., Roche, A.F., Johnson, C.L.: *CDC Growth Charts: United States*. *Advance Data* (December 2000)
12. Kay, S.M.: *Fundamentals of Statistical Processing*, 1st edn., vol. 1. Prentice-Hall, Englewood Cliffs (1993)
13. Szabo, T.L.: *Diagnostic Ultrasound Imaging*. Elsevier Academic Press, Burlington (2004)
14. Gallager, R.G.: *Low Density Parity-Check Codes*. MIT Press, Cambridge (1963)