TADAA: Towards Automated Detection of Anaesthetic Activity

B. R. Houliston¹; D. T. Parry¹; A. F. Merry²
¹AURA Laboratory, School of Computing & Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; ²Department of Anaesthesiology, Faculty of Medicine, University of Auckland, Auckland, New Zealand

Keywords
Anesthesiology, task performance and analysis, automated pattern recognition, radio frequency identification device

Methods: Active RFID tags were attached to anaesthetists and various objects in a high fidelity OR simulator, and anaesthetic procedures performed. The anaesthetists’ LOSs were calculated using received signal strength (RSS) measurements combined with machine learning tools including Self-Organizing Maps (SOMs). These LOSs were compared to those derived from video recordings.

Results: SOM clustering was effective at determining anaesthetists’ LOS from RSS data for each procedure. However cross-procedure comparison was less reliable, probably because of changes in the environment.

Conclusions: Active RFID tags provide potentially useful information on LOS at a low cost and with minimal impact on the work environment. Machine learning techniques may be employed to handle the variable nature of RFID’s radio signals. Work on mapping LOS data to activities will involve integration with other sensors and task analysis techniques.

1. Introduction

Recent years have seen an increased interest amongst researchers in recording and analysing common activities in a formal and objective way, in order to increase safety and reduce error. In safety-critical domains, such as aviation, such task analysis has identified the role of ‘human factors’ in accidents and near-misses [1], and informed the design, development, and evaluation of interventions to mitigate the risks. The medical community has begun to use techniques borrowed from such sources [2] to analyse activity, and the impact of human factors. Anaesthesia, one of the most critical and complex areas of medicine, has long been seen as a suitable and worthwhile field for the application of similar techniques [3]. A particularly important study in the New Zealand context found that human factors, such as poor record keeping, ignorance of standards or failure to adhere to them, and lack of communication, are major contributors to preventable adverse events during anaesthesia [4].

Weinger [5] states: “A scientific description of the anaesthesiologist’s task patterns and workload would aid in our understanding of the nature of the anaesthesiologist’s job and provide a more rational basis for improvements” in processes, equipment, OR layout, training and other aspects of anaesthesia. Having a ‘scientific description’ is particularly important in medicine, with its requirement for improvements to be evidence-based [6]. However, existing methods for anaesthetic task analysis are limited in their ability to produce scientific data in a timely, cost-effective way.

This project, Towards Automated Detection of Anaesthetic Activity (TADAA), is intended to identify methods of recording more scientific activity data, that are suitable for use in the OR. Ultimately such data, collected over many procedures, should allow a better understanding of anaesthetists’ activity, and inform the building of automated systems to detect and prevent error. The next section briefly reviews recent literature on anaesthetic task analysis and automated activity detection. The methods section describes the prototype TADAA system, including SOM analysis. Results from laboratory testing of RFID equipment are then presented, along with initial findings from SOM analysis of data captured during high fidelity simulated anaesthetic procedures. Finally, the paper concludes with outstanding issues and plans for future development.
1.1 Anaesthetic Task Analysis

During a procedure, the anaesthetist spends most of his or her time in a small area, typically 4–5 square metres, which we call the anaesthetic triangle. As shown in Figure 1, the vertices of the triangle are: 1) the drug trolley, where syringes and infusions are prepared; 2) the anaesthetic machine, containing monitors showing the patient’s vital signs and the controls for delivering anaesthetic gases; and 3) the patient.

Anaesthetists currently record some of their own activity in an anaesthetic record completed for each procedure. While this is a clinically important task, treating the patient always takes precedence. Thus parts of the record are often completed from memory, or with normalised activity [7], and its accuracy cannot be guaranteed. When more rigorous activity recording is required, the gold standard approach is to employ independent observers [5, 8]. The observers, either in the OR or viewing video recordings, classify anaesthetists’ activity into a number of a priori categories. A typical list of categories is shown in Table 1.

Human observers are not the ideal instrument for recording scientific data. They may make errors when fatigued, distracted, or their view is obstructed. Maintaining intra- and inter-observer consistency in classifying activities can be difficult [9]. The cognitive limitations of observers, and the user interface of typical recording tools, necessarily limit the level of detail that can be recorded [5]. Observers may distract OR staff, potentially causing changes in behaviour that confound the research or even posing a safety hazard. Using video recordings can mitigate some of these problems, but a camera's resolution and field-of-view tends to be more limited, and recording frequently raises privacy concerns for staff and patients [10]. Combined with the high cost of employing observers, these factors make task analysis impractical for ongoing use. Most published studies using observers cover fewer than 50 procedures.

An alternative to observation is a theoretical approach, such as the hierarchical task analysis (HTA) of anaesthesia recently performed by Phipps et al. [11]. Based primarily on a literature review and interviews with experts, the three stages of an anaesthetic procedure — Induction, Maintenance, and Emergence — were decomposed into their component activities. Each of those activities was then decomposed further, until the lowest level activities were reached. While HTA produced a more detailed list of activities (196, compared to the 29 in Table 1), it is arguably subjective. For example, the decomposition of ’Prepare drugs’ in [11] differs markedly from another HTA for drug preparation [12]. The activity documented in an HTA also tends to be normalised, reflecting standard practice rather than actual practice.

Thus both observation and theoretical task analysis seem poorly suited to capturing scientific — detailed, accurate, and objective — data. Attention has recently turned to automated methods of activity detection and analysis.

1.2 Automated Activity Detection

Activity detection in anaesthesia has traditionally focused on the patient. Computerised systems that record patients’ vital signs, and raise alarms when certain conditions occur, are widespread [7]. However systems to record anaesthetists’ activity are rare, with none seeming to have progressed beyond very limited use. Examples include general activity recognition from analysis of video [10] and/or audio [13], intubation...
assessments using RFID-tagged paraphernalia [14, 15] and detecting drug administration based on changes in patient vital signs [16] or scanning barcoded syringes [17].

A similarly small pool of automated activity detection systems used in the wider hospital environment has been documented, including [18–21]. There is a much larger body of research reporting on systems to recognize everyday activities in home and office environments, including [22–26].

Activity recognition systems can be characterised in various ways, including the type of sensor(s) used. In addition to those mentioned above, common choices are accelerometers [22], motion sensors and pressure plates [24]. Systems can be characterized by whether sensors are part of the environment [23, 24], or worn or carried by the research subject(s) [22, 26], and by whether the sensors are detecting body motion (such as gestures) [10, 22], location or movement around an environment [23, 24], or object use [14, 26]. Sensor data can be analyzed for activity using a priori rules [23], or machine learning [22, 26].

1.3 Location, Orientation and Stance

Location, orientation and stance (LOS) characterize the physical orientation or posture of individuals. Some activities are strongly related to a particular LOS. For example, intubation can only be performed when standing at and facing the patient. However LOS is not able to unambiguously identify every activity. For example, updating the anesthetic record can be performed anywhere in the OR.

LOS measurement supports activity detection in three ways. Firstly the LOS can be used alone or in combination with other data to identify potential activity. Previous work [27] has demonstrated that LOS can help to identify objects with which people are interacting. Secondly LOS can be used to cue other sensors – for example video or audio sensors. Thirdly LOS can be used to exclude potential activities. For example, an anesthetist cannot be adjusting the anesthetic machine if he or she standing at the drug trolley.

It was observed that these three parameters are strongly associated with each other for particular activities, and can distinguish between them. Thus, grouping LOS combinations allowed greater certainty of identification of an activity than assessment of each element separately.

2. Methods

2.1 TADAA System Design

In selecting a sensor technology, environmental sensors are preferred to worn or carried sensors, to minimize hygiene risks and the possibility of distracting anesthetists or impeding their work. RFID was selected as the most suitable sensor technology, as it offers both identification and movement sensing functionality. RFID is conceptually similar to barcode technology, using scanners (‘readers’ in RFID nomenclature) and labels (‘tags’) attached to people or objects of interest. The major difference between the two technologies is that RFID tags wirelessly transmit their ID numbers to the reader, rather than being optically scanned. A very good introduction to RFID technology is given by Want [28].

RFID is well represented in the activity recognition literature, and is becoming increasingly common in hospitals for a range of applications [29]. It does have disadvantages. RFID data tends to be both noisy and incomplete [30]. There is also some evidence of electromagnetic interference with electronic medical devices common in the OR [31], although this appears to be manageable in the case of low-power readers and when care is taken with placement [32, 33].

Activity detection with RFID is typically based on object use. For anesthesia that has required tagging of items such as drug ampoules and syringes [34] or intubation equipment [14, 15]. Even with the cheapest RFID tags, the cost and preparation time necessary to tag many single-use items may be prohibitive. For items that are re-used, tags that can withstand medical-grade sterilization [35–37] are available, but they are significantly more expensive.

The TADAA system instead uses RFID to perform location sensing of individuals. An RFID-tagged person can be located by readers embedded in the environment, using the received signal strength (RSS) of their tag’s transmissions. The RSS varies by the distance between tag and reader. Thus for a given RSS, distance can calculated using the principle of lateration-by-attenuation (LAT) [38]. Although lateration via signal strength is a widely used method for location sensing [39], its use is complicated by the fact that RSS is affected by many factors other than distance. Tag orientation, intervening objects or people, and atmospheric conditions, all impact on RSS in ways that are difficult to quantify. The OR presents additional challenges, such as several other electronic devices and many metal surfaces, some of which are movable.

The resulting multipath reflection, attenuation, electromagnetic interference, and far near field effects mean that the relation between RSS and distance deviates significantly from the inverse square curve that would be expected in free space. Real-time location systems (RTLSs), a common RFID application in hospitals, employ LAT to locate tagged people or equipment in wards, floors or entire buildings. However, such systems are generally only accurate to within a few metres, too large a margin for the TADAA system.

Research suggests that more precise location is possible with LAT over smaller areas [40, 41]. However, that research has either used very simple environments (40) and/or supplemented LAT with additional data [41].

Three sets of RFID tags were used in testing the TADAA system. Firstly, the anesthetists wore two tags, one on the front of their scrubs and one at the back. Secondly, seven tags were placed at ‘landmark’ locations within the anesthetic triangle, including the drug trolley, anesthetic machine, and the patient. This approach was used by the LANDMARC system [41] to increase location sensing accuracy by comparing theRSSs of target tags with those of such landmark tags at known locations. Thirdly, six tags were suspended from the ceiling to give a baseline measurement of any environmental changes that may also affect other tags. Readers were attached to the ceiling at three locations, outside the bounds of the anesthetic triangle.
Anticipating that LAT would be insufficient for LOS sensing, an empirical approach to deriving LOS from RSS values was also tested, in the form of Self-Organising Maps (SOMs) [42]. SOMs are a type of neural network that use unsupervised learning to identify clusters in data from multiple inputs. SOMs are a common exploratory tool, well suited to the highly complex interactions of the various factors that influence RSS, and can be seen as a way of reducing the dimensionality of datasets [43]. TADAA used SOMs to map RSS data—collected into ‘fingerprints’ each with 45 points (15 tags × 3 readers) – to LOSs. The maps produced by SOM tools provide a highly intuitive visualization of clusters, with their physical proximity on the maps reflecting the similarity of the clusters.

The TADAA evaluation comprised two stages; firstly confirming that the selected RFID equipment is suitable for LOS sensing, and secondly confirming that the system can provide accurate LOS sensing in the OR, at less cost and risk of distraction than a human observer.

2.2 RFID Equipment

The TADAA system used active RFID equipment manufactured by Wavetrend, operating in the 433 MHz band. Tags transmit every 1.5 seconds. A small random variation is built in to reduce tag collision, but if multiple tags transmit at the same instant, then all but one transmission will be lost. The reader ID, tag ID, date, time, and an RSS value are captured for each transmission. Three readers were used, so that for each tag transmission, up to three RSS values were received.

LAT accuracy largely rests on three aspects of RSS: granularity, i.e. – that each RSS value corresponds to a small range of distances; stability, i.e. – that RSS doesn’t change when the tag isn’t moving; and predictability, i.e. – that RSS changes predictably when a tag is moving.

Effective activity recognition requires that the time between tag readings should be less than the shortest activity duration of the activities being performed [40, 41]. The only published research to discuss the required time between sensor readings for anaesthetic activity recognition suggests that “the anaesthetist is involved in virtually no activities completed in less than two seconds” [44]. Data collected during previous manual observations by the authors of this paper found that 5–6% of activities had a duration of two seconds or less. Thus two seconds was adopted as a desirable time between reads.

Tests were performed in the AURA Laboratory to determine how well the RFID equipment satisfied these requirements. The tests included the effects on RSS of known causes of variation, such as having an anaesthetist wear the tag, change tag-reader orientation by changing stance, or obstruct the tag by standing between it and the reader.

2.3 LOS Sensing

The TADAA system was deployed in a high-fidelity OR simulator. This is an area set up with standard OR equipment, but used for training with mannequins replacing the patient and a large number of video cameras able to record events. Data was first collected from two dry runs during which one of the authors simply adopted 17 typical LOSs, without performing any activity.

The main data collection was from 40 simulated procedures, performed over 20
days by 20 different anaesthetists. Along with the RFID data, the procedures were also videotaped and recorded by a human observer. Each anaesthetist was asked to rate the distance caused by the RFID tags, the RFID readers, and the observer on a visual analog scale (VAS).

The videos served as the ground truth, with the anaesthetist’s LOS determined manually for each second of video. The RFID data was aggregated to produce one record per second, combining the 45 RSS values from all 15 tags and three readers. Missing RSS values (due to the 1.5 s transmission rate and lost transmissions) were filled in with the average RSS for the relevant tag, reader, and known RSS values.

To perform SOM clustering, the RFID data was then processed using the SOM functions in MATLAB’s Neural Network toolbox (MATLAB). Analysis was first performed on the dry run data, with SOMs trained on the data from dry run 1 and tested on the data from dry run 2, using different combinations of SOM size and tag subsets.

SOMs were evaluated using precision, or positive predictive value. Each cluster was assigned to the LOS which contributed the largest number of records. Those records were considered true positives (TP). Any other records in the cluster, from other LOSs, were considered false positives (FP). The SOM’s precision was calculated as total TP/(total TP + total FP).

Results from the analysis of the dry runs informed further analysis of data from the simulated procedures.

3. Results

3.1 RFID Performance

The average time between tag reads for 15 tags was approximately 1.98 s, with a standard deviation of 1.2 s. The minimum RSS value recorded during testing with the tag-reader distance up to distances of 6 m was 89 and the maximum 212. RSS granularity was therefore 123 units over 6 m, or just under 5 cm per unit.

When tags were stationary at 2 m from the reader, the standard deviation in RSS was approximately 0.8% of the average RSS value. However, each tag-reader combination had a different average RSS.

When tags were moved RSS did vary by distance for all tags, but not in a predictable way. The RSS-distance relations shown in Figure 2 are typical: an overall downward trend, but with peaks and troughs. Different tag-reader combinations exhibited different RSS-distance relations. As Figure 2 illustrates the RSS at a given distance varied by up to 40 units depending on whether the tag was moving away from or towards the reader. Table 2 shows the maximum effect on RSS of common interactions between anaesthetists and tags.

3.2 LOS Sensing

Table 3 shows the maximum precision achieved by training SOMs on the data from one procedure. When the most precise SOM was tested on data from another procedure, rather than the 17 used in the dry runs, other data for all procedures on the same day, or other data for all procedures. For the sake of completeness we also tried no data cleaning.

● Larger SOM sizes were tried, since video analysis identified 150–200 LOSs per procedure, rather than the 17 used in the dry runs.

Table 4 shows the maximum precision achieved by training SOMs on the data from one procedure. When the most precise SOM was tested on data from another procedure on the same day, precision decreased to 54%. When the SOM was tested on data from procedures on different days, precision fell within the 40%–50% range. Similar results were found when repeating the SOM analysis starting with different procedures.

The anaesthetists’ VASs for distraction were converted to values in the range 0 (No distraction) to 100 (Worst distraction). Table 5 shows a summary of results.
Table 4 Maximum SOM precision for procedure 39 data

<table>
<thead>
<tr>
<th>Cleaning</th>
<th>SOM Size 10</th>
<th>SOM Size 20</th>
<th>SOM Size 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>17%</td>
<td>40%</td>
<td>77%</td>
</tr>
<tr>
<td>Using data from same procedure only</td>
<td>58%</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>Using data from same day only</td>
<td>56%</td>
<td>84%</td>
<td>92%</td>
</tr>
<tr>
<td>Using all data</td>
<td>39%</td>
<td>61%</td>
<td>78%</td>
</tr>
</tbody>
</table>

Table 5 Distraction ratings (n = 20)

<table>
<thead>
<tr>
<th>Distraction caused by</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags</td>
<td>0.25</td>
<td>25.5</td>
<td>5.55</td>
</tr>
<tr>
<td>Readers</td>
<td>0.75</td>
<td>10.75</td>
<td>3.94</td>
</tr>
<tr>
<td>Observer</td>
<td>1.00</td>
<td>70.00</td>
<td>18.25</td>
</tr>
</tbody>
</table>

4. Discussion

Based on the laboratory test results, active RFID equipment appears suitable for LOS sensing in most respects. The average time between tag readings of 1.98s is just under the desirable 2-second value, and certainly a good deal better than previous results (7.5 s – 30 s) reported in [40, 41].

Likewise, RSS granularity of 123 units (@ approximately 5 cm) is much better than the 8 units (@ approximately 30 cm) reported in [40]. The RSS standard deviation is small in relation to the average, indicating that RSS is very stable. The major challenge presented by the RFID equipment is the RSS-distance predictability. Although it was not expected that the tags would exhibit ideal inverse square relations [38], it was surprising to observe the variation between different tag-reader combinations. This would prevent LANDMARC-style RSS comparisons across tags [41]. However this can be regarded as just another RSS variation, alongside movement, orientation, obstruction by OR staff, and EMI. The multiple sources of variation and their complex interaction, as shown in Table 2, suggest that a basic formula would not have performed well for location sensing in this environment, and that a clustering tool such as SOMs was a better starting point.

The assumption that the RSSs of the ceiling tags would remain relatively stable was proven to be true only for the tags closest to each reader. For the ceiling tags directly over the anaesthetic triangle, RSS does appear to be affected by anaesthetists’ movement. It was surprising to see that, at larger SOM sizes, the landmark and ceiling tags both seemed to be more useful for LOS sensing than the anaesthetist tags (Table 3).

The very high precision scores of SOMs trained on the dry run data (Table 3) were achieved in a very simple environment, with only one person in the work area and a small number of LOSs. However, greater than 90% precision was also achieved in the more realistic simulated procedures. Such high precision would be unlikely in a real-time system, since it relied on having complete data for the period being analysed. Having to fill in missing RSS values using data from previous days, or even earlier procedures in the same day, can reduce precision significantly, although it is still an improvement over no cleaning at all (Table 4).

Some decrease in precision when testing SOMs across procedures/days was expected, but such large decreases were surprising. Comparing the RFID data for matching tag-reader-LOS combinations across the two dry runs revealed a large variation in RSS averaged across all the tags, suggesting that some general environmental effect may have accounted for some of the difference. The remainder of the difference may reflect changes in the location and orientation of anaesthetic and landmark tags between procedures. For example, the operating table is frequently moved between procedures so that the floor underneath can be cleaned, and even a small location change in repositioning it could result in large differences in RSS (Figure 2 and Table 2). These differences may be detected, and corrected for with a calibration process before each procedure or at the start of each day.

Unsurprisingly, the anaesthetists found the RFID tags and readers much less distracting than a human observer (Table 5). The cost of the TADAA system’s hardware is approximately the same as a human observer’s tablet PC. The system’s ongoing cost, replacing active RFID tags every five years, is significantly less than the long term costs of an observer.

5. Future Work

We continue to develop the TADAA system’s RFID data cleaning, SOM analysis, and HMM analysis algorithms for activity recognition testing.

In the medium term, we intend to complement RFID with additional sensors. RFID is best suited to LOS-specific manual activities (Table 1). Detecting LOS-independent activities would be better done by other sensors. Recording communication behaviour, for example, might use a microphone suspended above the anaesthetic triangle. We are also considering means of recognising not just when a drug is administered, but which drug and how much. Analysis of patient vital signs may assist this, as in [16]. In the future we wish to confirm that, given accurate LOS data, the system can recognise activity as well as or better than a human observer.

In order to recognize activities from a time-series of LOSs it is planned to use a Hidden Markov Model (HMM) [45], reflecting the largely sequential and ordered nature of activity during the Induction and Emergence phases of an anaesthetic procedure. In addition to the LOS data, the HMM would be informed by an anaesthesia HTA adapted from [11], and by data on activity frequencies and duration extracted from previous observation. HMM development is ongoing, and not covered in the results presented in this paper.

We ultimately see the TADAA system as a tool for improving patient safety. Activity data collected over time could be mined to produce ‘norms’ for procedure types, patient conditions, and so on. These would be as detailed as the HTA in [11] but based on
actual practice. Anesthetic activity detection in real-time could then compare actual activity with the norm, recognize deviations and raise alarms.

6. Conclusion

Task analysis has the potential to improve patient safety. But to do so it must capture accurate data to satisfy the requirements of evidence-based medicine, and not be too intrusive or expensive to implement in practice. Current task analysis methods, based on theory or observation, are unlikely to meet these criteria.

The TADAA system is intended to automate anaesthetic task analysis, through use of active RFID technology, and a combination of SOM and HMM machine learning informed by existing task analysis data. Testing in our lab and in high-fidelity OR simulations suggest that the TADAA system is less intrusive and expensive than observers. However, accurate location sensing, a first step towards activity detection, remains a challenge in the face of a complex interaction of factors affecting RFID’s radio signals.

As development of TADAA continues we envisage adding new sensors, and means of storing, mining and visualising activity data. We see the system as potentially a valuable tool to improve patient safety in anaesthesia, and in other medical disciplines.

Acknowledgments

The authors wish to thank Dr Craig Webster and staff at the University of Auckland’s Advanced Clinical Skills Centre, particularly Dr Jane Torrie and Kaylene Henderson, for allowing us to participate in the simulated anaesthetic procedures, and to HamIT Ltd for supply of RFID equipment.

References