Proceedings of the Twenty-Second Innovative Applications of Artificial Intelligence Conference (IAAI-10)

Reinforcement Learning for Closed-Loop Propofol Anesthesia: A Human Volunteer Study

Brett L. Moore, MS[†] and **Periklis Panousis, MD**[‡] and **Vivek Kulkarni, MD, PhD**[‡]
Larry D. Pyeatt, PhD[†] and **Anthony G. Doufas, MD, PhD**[‡]

[†]Department of Computer Science, Texas Tech University, 302 Pine St, Abilene, TX, 79601
[‡]Department of Anesthesia, Stanford University School of Medicine, 300 Pasteur Dr, Stanford, CA, 94304

Abstract

Research has demonstrated the efficacy of closed-loop control of anesthesia using the bispectral index (BIS) of the electroencephalogram as the controlled variable, and the development of model-based, patient-adaptive systems has considerably improved anesthetic control. To further explore the use of model-based control in anesthesia, we investigated the application of reinforcement learning (RL) in the delivery of patient-specific, propofol-induced hypnosis in human volunteers. When compared to published performance metrics, RL control demonstrated accuracy and stability, indicating that further, more rigorous clinical study is warranted.

When compared to population-based dosing, patient-specific drug administration is generally preferred in the clinical practice of anesthesia. Computer-controlled drug delivery systems have been investigated as a means of achieving patient-specific anesthesia, and their use is associated with a number of favorable patient outcomes, including decreased intraoperative drug consumption and shortened postoperative recovery times (Liu et al. 2006; Servin 1998; Theil et al. 1993). Historically, the application of proportional-integral-derivative (PID) control in closed-loop anesthesia has demonstrated moderate success (Absalom and Kenny 2003). However, success has been constrained by limitations in the control method, as well as the complexity of human physiology (Wood 1989). To improve control performance, clinical study has broadened to include techniques commonly associated with intelligent systems, most notably fuzzy control (De Smet et al. 2008; Esmaeili et al. 2008; Carregal et al. 2000; Schaublin et al. 1996).

Reinforcement learning (RL), another intelligent systems technique, has demonstrated proficiency in difficult robotic control tasks (Gullapalli 1993); however, RL has no reported application to clinical control problems, with the exception of work relating to this study (Moore et al. 2004). Despite the lack of clinical application, reinforcement learning is not far removed from medicine since its fundamental principles (dynamic programming and value function optimization) have been applied to depth of anesthesia control with favorable results (Hu, Lovejoy, and Shafer 1994).

Because RL's aptitude for specialized control tasks remains incompletely explored, the objective of this study was to investigate the application of reinforcement learning to closed-loop control of intravenous propofol anesthesia in healthy human volunteers. Accordingly, an RL agent was developed, tested *in silico*, and then evaluated in volunteers under an IRB-approved study protocol in the Department of Anesthesia at Stanford University School of Medicine.

Background

Propofol-Induced Hypnosis

Propofol is a short-acting sedative administered intravenously to achieve induction and maintenance of general anesthesia. Propofol suppresses cortical brain function (hypnosis) but offers no analgesic effect (pain relief).

The anesthesia community has studied automated delivery of propofol for two principal reasons. First, the short-acting nature of the drug (characterized by rapid onset and recovery) permits titration to effect. Second, indication of propofol effect may be observed in the electroencephalogram (EEG) (Glass et al. 1997). Currently, the bispectral index of the EEG, or BIS (Aspect Medical Systems, Newton, MA), enjoys the greatest clinical acceptance as a measure of hypnotic effect. BIS is a statistically derived indicator of cortical activity that lies in the range [0, 100] (Rampil 1997).

Motivation for Good Control of Hypnosis

BIS has been recently studied as a mitigation for reducing the risk of intraoperative awareness, defined as conscious behavior (motion, vocalization, etc.) during surgery or post-operative recall of intraoperative events (explicit or implicit). Unintentional intraoperative awareness can challenge the anesthetist because doses ensuring adequate hypnosis may be associated with hemodynamic and/or respiratory instabilities in sensitive patients (i.e. trauma, critically-ill, elderly, etc.). While the incidence of intraoperative awareness is estimated to be low (0.13%) (Sebel et al. 2004), it can be extremely traumatic for the patient. BIS monitoring has been recommended as a preventative measure (Sandin et al. 2000) and has been reported to reduce the incidence of unintended intraoperative awareness (Myles et al. 2000). This finding remains controversial since this evidence comes from observational clinical trials (Avidan et al. 2008), and

the execution of a convincing prospective clinical trial is logistically challenging.

At first glance, the risk of intraoperative awareness implies that "deeper is better"; however, higher doses of propofol are correlated with incidences of respiratory and hemodynamic depression. Emerging research corroborates a need for balance: Lindholm et al. report a possible causal link between deep anesthesia (BIS < 45) and postoperative morbidity (Lindholm et al. 2009). As before, this conclusion requires further substantiation before universal acceptance.

These opposing concerns (awareness versus toxicity), as well as the favorable outcomes cited above, associate good control of intraoperative anesthesia with good patient care. Consequently, closed-loop control of propofol-induced hypnosis is well-represented in the literature (Struys et al. 2007; 2004; Absalom and Kenny 2003; Leslie, Absalom, and Kenny 2002; Absalom, Sutcliffe, and Kenny 2002; Sakai et al. 2000; Struys et al. 2001), but accurate and stable control of intraoperative hypnosis remains an unsolved problem.

Challenges to Optimal Control of Hypnosis

Existing closed-loop drug delivery systems are typically based on population models of drug effect (Vuyk et al. 1995), making them ill-equipped for accurate drug delivery in the individual. Age, gender, and ethnicity, as well as disease and surgical intervention (Schnider et al. 1998; Barvais et al. 1996), are known to affect a patient's response to propofol infusion. Control of propofol-induced hypnosis is also complicated by delays in action and effect (*transport delay* in control literature). The delays are variable, hysteretic, and demonstrate flow rate dependence (Struys et al. 2007; Pilge et al. 2006). Furthermore, environmental conditions can influence patient response; for example, routine surgical events (incision, manipulation, etc) challenge good control (Röpcke et al. 2001; Ausems et al. 1986).

Reinforcement Learning

Reinforcement learning (RL) is intelligent control method that presents a structured, mathematically robust mechanism for goal-directed decision-making in which long-term gain is to be maximized (Sutton and Barto 1998). Unlike supervised learning methods, no examples of desired behavior are provided during training; instead, behavior is guided through positive or negative reinforcements. Knowledge is gained through experimentation: actions are chosen, effects are observed, and rewards are gained accordingly.

Methods

A study protocol was designed to evaluate the reinforcement learning (RL) agent using widely accepted measures of closed-loop control performance. The agent was originally developed in the Department of Computer Science at Texas Tech University. Several simulation trials, as well as preliminary studies in human volunteers at Stanford University, were used to improve the agent's performance and produce a clinically suitable controller. The final clinical evaluation, conducted in the Department of Anesthesia at Stanford University School of Medicine, consisted of fifteen, consecutively-studied, healthy volunteers.

Agent Architecture

Agent Inputs To achieve and maintain a desired level of hypnosis ($\mathrm{BIS}_{\mathrm{target}}$), the agent first observed the volunteer's bispectral index ($\mathrm{BIS}_{\mathrm{measured}}$) on five-second intervals, as reported by an A-2000 BIS monitor (Aspect Medical Systems, Norwood, MA). Since BIS is an inherently noisy signal, $\mathrm{BIS}_{\mathrm{measured}}$ was smoothed using a low-pass filter. Two control inputs were then computed: E and ΔE. E was defined as $(\mathrm{BIS}_{\mathrm{smoothed}} - \mathrm{BIS}_{\mathrm{target}})$, and ΔE was defined as the change in E over 15s, or $(E_t - E_{t-2})$.

In pilot studies, the combined effects of BIS measurement noise and transport delay resulted in oscillatory behavior. These confounding influences were successfully mitigated by conditioning E and ΔE with sets of fuzzy membership functions (Zadeh 1965). The fuzzy set membership for E and ΔE was assessed using two sets of triangular membership functions, $\mu_N(x)$, $\mu_Z(x)$, and $\mu_P(x)$. The resulting six-dimensional feature vector served as the agent's input.

Agent Actions The agent delivered propofol via a catheter placed in the antecubital vein (elbow) using a precision syringe pump (Pump 33, Harvard Apparatus, Holliston, MA). During control, the agent could select an infusion rate from a discrete set ranging from 0.0–6.0 ml/min. Once a rate was selected, the action remained in effect for five seconds. The agent lacked the ability to directly reduce propofol concentrations in the patient, resulting in a condition of asymmetric control.

Knowledge Representation During the control process, the RL agent is expected to observe the patient's state, and then select the appropriate propofol dose from its control policy. To learn the optimal control policy, the agent accumulated its experience in *value functions*, or mathematical descriptions of the utility of patient states and infusion rate selections. In RL, learning is accomplished through iterative function approximation, and value functions must be represented in suitable form; tables, decision trees, neural networks, and weighted polynomials have precedence in the literature. Of these, the uniformly discretized table is favored for its ease of implementation and mathematical robustness (Baird 1995).

In this study, the agent's state observation vector consisted of the six fuzzy membership values described previously. The tabular value function approximator bounded each feature element to the range [0, 1], and then partitioned the range into ten discrete bins. Thus, a table of 10^6 entries was associated with each discrete infusion rate.

Reward The agent's objective was to achieve and maintain the selected BIS target (i.e. minimize the absolute control error $|E|$). Since the task lacked an explicit terminal state, the reward function $r_{t+1} = -|E_t|$ was chosen.

Agent Training

Because the naive, uninformed agent can make arbitrarily poor dosing decisions, a simulated intraoperative patient was

developed for agent training. This *in silico* patient also presented an advantage in its rapid simulation of hypnotic episodes. Reinforcement learning is fundamentally a process of statistical estimation, and a large number of training episodes were needed to achieve clinical readiness.

The principle role of this virtual patient was to model the time-dependent effects of propofol infusion, collectively known as the *pharmacokinetic* and *pharmacodynamic* (PK/PD) responses to propofol. (A drug's pharmacokinetic properties describe its distribution within the body, and pharmacodynamic attributes characterize the dose effect.)

Modeling Propofol Effect Propofol pharmacokinetics were modeled using Schnider's three-compartment model (Schnider et al. 1998), which provides the central, rapid, and slow compartments to estimate the time-dependent distribution of propofol. In this model, propofol is introduced into the central compartment via intravenous infusion; after infusion, the drug is free to interact with the rapid and slow compartments through first-order, gradient driven flow. The rapid and slow compartments represent collections of tissues with high and low propofol transport coefficients; however, it is important to note that the model was derived from empirical observations. As such, these compartments may not have a direct, obvious mapping to physiological systems.

Figure 1 illustrates the model and its transport coefficients, which vary with patient height, weight, gender, and age. As shown, the coefficients are subscripted to indicate direction of flow (from, to) because the coefficients may differ directionally, i.e. the central-to-slow coefficient (k_{cs}) is not equal to the slow-to-central coefficient (k_{sc}). Metabolic losses of propofol are represented in k_{c0}.

Because an infusion of propofol exhibits a 2.7 minute time-to-peak effect in BIS (Schnider et al. 1998), a fourth *effect site* compartment was implemented. The resulting transport coefficient, k_{e0}, was assigned a value of 0.17 (Doufas et al. 2004). The effect-site compartment is assumed to possess negligible volume; hence, effect-site interaction is assumed to be one-way.

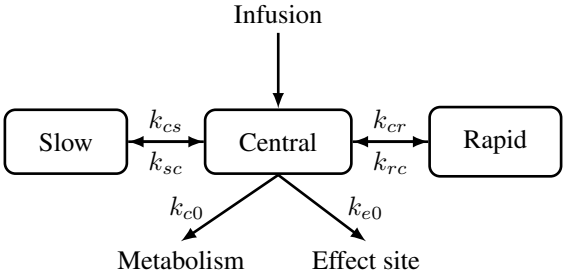

Figure 1: Schnider's pharmacokinetic model of propofol. The drug is infused into the central compartment, and concentration gradients govern the subsequent transport to other compartments.

To model the hypnotic effect of propofol, a nonlinear pharmacodynamic model was developed using previously obtained data (Doufas et al. 2004). A three-layer perceptron network was trained to associate arterial concentrations of propofol with observed BIS, allowing the agent to generally predict propofol effect from estimated effect site concentration. Figure 2 illustrates the observations of BIS and propofol concentration, as well as the median fit.

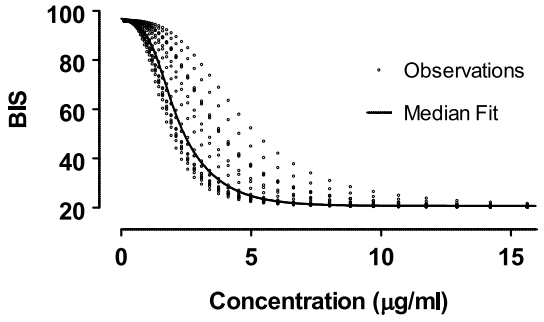

Figure 2: The Doufas nonlinear model for propofol pharmacodynamics. Doufas et al. observed the propofol/BIS response in eighteen young, healthy subjects. To model propofol pharmacodynamic effect in this study, a neural network function approximator was used to fit the median dose curve (highlighted here).

Learning Algorithm Watkins' Q-learning algorithm was used to train the agent (Watkins 1989). Q-learning is a temporal difference learning method characterized by model-free, off-policy learning. The algorithm is mathematically robust (Tsitsiklas and Van Roy 1996; Dayan 1992), a property that has contributed to the method's popularity in applied RL applications. Furthermore, this method has been shown to work well with tabular function approximators. To speed learning, the Q-learning algorithm was implemented with eligibility traces ($\lambda = 0.8$) (Sutton and Barto 1998).

Training Agent training consisted of a sequence of simulated hypnosis episodes using a standardized intraoperative patient prototype (male, 21 yr, 170 cm, 75 kg). To aid in learning a general association of propofol infusion and patient response, the patient's k_{e0} was randomly selected $[0.17 \pm 25\%]$ at the beginning of each episode. This measure, of which the agent remained unaware, influenced the timing and magnitude of peak BIS effect as illustrated in Figure 3. In this figure, the effect of k_{e0} variation in a simulated patient is shown with a bolus of propofol delivered at $t = 0$ min. The bolus was allowed to distribute according to the Schnider pharmacokinetic model under the selected k_{e0}. As shown, a larger k_{e0} represented a more "tightly coupled" system in which propofol was transported to the effect site more readily, yielding a deeper level of hypnosis. For emphasis, Figure 3 highlights the minimum hypnotic levels, as well as the times of their occurrence. While the time of peak effect varied by approximately 25 seconds, the range in peak effect varied by more than 20 BIS points.

To ensure adequate exploration of the state-action space, each episode began with an *exploring start* in which a BIS target was randomly selected, and random propofol quantities were assigned to the three major PK compartments. The

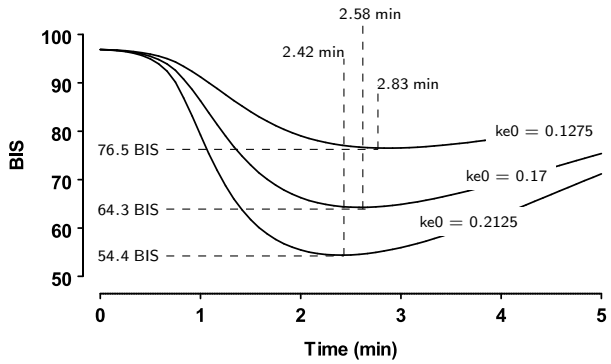

Figure 3: Pharmacodynamic effect of variation in k_{e0}. To demonstrate the individual subject variation associated with changes in k_{e0}, a bolus of propofol was delivered to a simulated patient and propofol distribution was modeled over time. For comparison, k_{e0} was selected at 0.17, 0.1275 ($0.17 - 25\%$), and 0.2125 ($0.17 + 25\%$). The points of peak BIS effect and their associated times are highlighted.

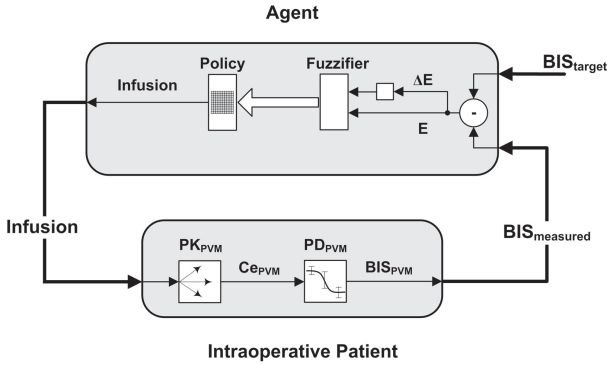

Figure 4: Interaction between the agent and simulated intraoperative patient. To estimate the agent's performance in clinical application, simulated episodes of hypnosis were performed on a population of 1,000 randomly selected patients. A *Patient Variability Module* (PVM) challenged the agent with an array of static and time-varying parameters, including variations in patient demographics, propofol sensitivity, and intraoperative events. PVM influence was not directly observable by the agent.

agent was then permitted to interact with the patient and accumulate reinforcements. After 1,000 action choices (5,000 simulated seconds), the episode was concluded, and a new one begun.

Training began with a step-size parameter $\alpha = 0.2$, horizon parameter $\gamma = 0.69$, and an exploration parameter $\epsilon = 0.01$. To assess the progress of learning, the sum of squared difference (SSD) was computed between intermediate control polices. When the SSD metric fell below a small threshold, α was halved, and learning resumed. This procedure continued until $\alpha = 10^{-5}$. In total, training required 5×10^7 episodes and approximately one week of CPU time on a contemporary desktop computer.

In silico Control Policy Evaluation

Prior to clinical application, the agent was evaluated in simulation. To assess the fitness of the candidate control policy, a Patient Variability Model (PVM) was constructed to challenge the controller with individualized patients (Figure 4). The PVM modeled both PK and PD variation: the PK_{PVM} component modeled changes in k_{e0} as described previously (see Figure 3), while the PD_{PVM} block modeled changes in propofol sensitivity (ΔBIS_{PVM}) as a sum of time-dependent and time-independent parameters (Moore, Pyeatt, and Doufas 2009). PVM influence can be summarized as:

$$(1) \quad BIS_{measured}(t) = BIS_{ideal}(t) + \Delta BIS_{PVM}(t)$$

Evaluation consisted of a sequence of sedation episodes in a population of 1,000 simulated patients possessing randomized demographic and PVM parameters. In each episode, the agent was tasked with achieving and maintaining propofol-induced hypnosis for 240 minutes. BIS targets were randomly selected (without replacement) from the set $\{40,50,60\}$. Once selected, a target remained in effect for 80 minutes.

Performance Evaluation

The steady-state control performance was evaluated using the four metrics of Varvel et al. (Varvel, Donoho, and Shafer 1992), which comprise the standard performance measures in closed-loop infusion control. These metrics build upon the instantaneous performance error (PE):

$$(2) \quad PE = \frac{BIS_{measured} - BIS_{target}}{BIS_{target}} \cdot 100$$

The median performance error (MDPE) indicates the control bias observed in a single patient and is computed as:

$$(3) \quad MDPE_i = \text{median}\,(PE_{ij}) \qquad j = 1 \ldots N$$

where i identifies a subject, and j iterates over the set of PE measurements for a subject. Median absolute performance (MDAPE) error reflects the accuracy of the controller in a subject:

$$(4) \quad MDAPE_i = \text{median}\,(|PE_{ij}|) \qquad j = 1 \ldots N$$

Wobble measures the intra-subject variability in performance error:

$$(5) \quad Wobble_i = \text{median}\,(|PE_{ij} - MDPE_i|)\ j = 1 \ldots N$$

Divergence is defined as the slope of the regression line computed through the observed MDAPE measurements. Positive values indicate an increasing difference in measured and target values; a negative divergence indicates more stable control.

In addition to the Varvel metrics, recent studies have computed the *Controlled* metric, the percentage of measurements in which the observed BIS was within $\pm$ 10 BIS (Struys et al. 2004) or $\pm$ 5 BIS (De Smet et al. 2008) of target. As an additional performance comparator, this study also reports the root-mean-square error (RMSE) computed for each maintenance control interval.

Table 1: Human Volunteer Demographics

Height (cm)	Weight (kg)	BMI (kg/m^2)	Age (yr)
174.5±9.6	72.2±10.0	22.0±1.6	20.7±2.5
$N = 15$		$(N_{male} = 11, N_{female} = 4)$	

Acceptance Criteria The literature does not provide a definitive guideline for clinically suitable control of propofol-induced hypnosis, but a survey of three contemporary studies (De Smet et al. 2008; Struys et al. 2004; Absalom and Kenny 2003) provides some reasonable performance goals: MDPE $\leq \pm 5.0\%$, MDAPE $\leq 7.5\%$, Wobble $\leq 5.0\%$, Divergence $\leq \pm 0.1\%$/hr, Controlled $\geq 80\%$, and RMSE ≤ 5.0 BIS.

Clinical Application After IRB approval (Stanford School of Medicine) and informed consent, we studied fifteen consecutively recruited healthy (BMI $\leq$ 25 kg/m^2, 18-45 yr) volunteers in the operating room at Stanford University Medical Center.

Volunteers fasted for at least six hours prior to the study and their vital signs were monitored according to the standards of the American Society of Anesthesiologists (ASA). After placement of the monitors (including BIS) an intravenous catheter was inserted at the elbow. The study began when the anesthesiologist directed the RL agent to achieve a randomly selected initial target (40 or 60). Once $\mathrm{BIS}_{\mathrm{target}}$ was achieved, the agent was permitted to regulate the level of hypnosis undisturbed for 15 minutes. A mild tetanic stimulus was then administered to the volunteer's thigh to simulate a destabilizing surgical event, and control was allowed to continue for an additional 15 minutes. At that time, the agent was directed to achieve the second $\mathrm{BIS}_{\mathrm{target}}$. Once the volunteer had stabilized at the second target, a similar procedure of maintenance and stimulus followed; then, the volunteer was then allowed to recover normally from sedation.

Analysis To analyze the steady-state control performance, automated tools identified and discarded the intervals of induction and target change. The resulting maintenance intervals were then scored using the methods applied in the *in silico* performance analysis (Equations 2–5). The $\mathrm{BIS}_{\mathrm{target}} = 40$ and $\mathrm{BIS}_{\mathrm{target}} = 60$ control periods were evaluated independently, then in aggregate form.

Results

Fifteen healthy human volunteers (11 males and 4 females) were recruited to assess the effectiveness of RL control in closed-loop delivery of intravenous propofol anesthesia (Table 1). Table 2 summarizes the agent's observed control performance in delivering propofol to achieve targets of $\mathrm{BIS}_{\mathrm{target}} = 40$ and 60. As shown, the average aggregate control metrics were within the desired performance values.

Table 2: Control Results

	$\mathrm{BIS}_{\mathrm{target}}$ 40	$\mathrm{BIS}_{\mathrm{target}}$ 60	Aggregate
Duration†	30.2±5.2	30.1±2.5	60.3±5.1
MDPE‡	1.0±5.6	−0.2±1.2	0.4±3.0
MDAPE‡	7.4±3.5	2.8±1.2	5.1±1.7
Wobble‡	6.2±2.6	2.6±1.2	4.5±1.5
Divergence*	< 0.001	< 0.001	< 0.001
RMSE§	4.5±1.7	2.9±1.1	3.7±0.9
Controlled‡	75.6±19.5	90.5±11.0	82.9±9.6

mean ± std. dev. †(min), ‡(%), *(%/hr), §(BIS)

Discussion

Figure 5 illustrates one promising aspect of RL control: patient-specific hypnosis. During each study, the data collection system computed the predicted bispectral index as the agent controlled the volunteer's level of hypnosis. Using the volunteer's demographic data, the agent's action history, and the Schnider-Doufas PK/PD model, an estimate of propofol effect was computed on five-second intervals. By comparing predicted and observed BIS values (Figure 5), the RL agent's ability to compensate for model mis-specification is evident. As shown, Volunteer A demonstrated an apparent sensitivity to propofol: the observed hypnosis level was consistently below the predicted value for most of the 30-minute period shown. Likewise, the RL agent compensated for an apparent propofol tolerance in Volunteer B. In this 30-minute period, the observed BIS was consistently above the predicted value, indicating that the volunteer required more propofol than predicted. These observations suggest that the reinforcement learning process yielded a patient-specific control policy that may be applied to a general population of volunteers.

Limitations

The principle limitation of this study lies in its controlled nature: the human volunteers were healthy and resembled those populations from which the PK/PD models were derived. Although the agent was challenged with intra- and inter-subject variation, it did not experience the full rigor of the intraoperative environment. Consequently, these results may be extrapolated to surgical patients with limited fidelity.

Future Directions

Given the favorable performance in both simulation and healthy human volunteers, it is reasonable to evaluate the agent in intraoperative patients to further assess the clinical utility of RL control. A study of agent performance under the rigors of the surgical environment and varying conditions of patient health should provide additional insight in the suitability of the technique. Likewise, it would be interesting to further explore the agent's capacity for patient-specific hypnosis with evaluations in patient populations lacking good

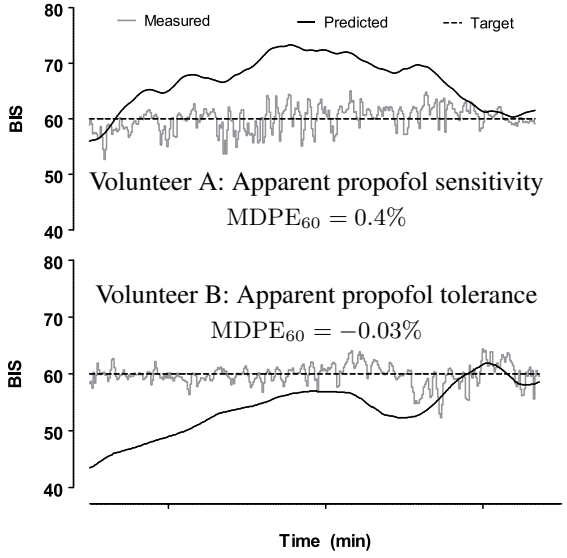

Figure 5: Examples of patient-specific anesthesia. Although the RL agent was trained using a standardized patient prototype, the agent demonstrated good control in subjects demonstrating apparent propofol sensitivity (Volunteer A) and apparent propofol tolerance (Volunteer B).

PK/PD models (critically ill, morbidly obese, etc).

This study has also identified some areas of improvement. First, the agent controlled hypnosis better at $\mathrm{BIS}_{\mathrm{target}} = 60$ (as indicated by most metrics). Since the agent's action choices were recorded during this study, it is anticipated that an improved agent can be developed by incorporating the resultant volunteer responses into the training regimen.

Second, the agent's reward structure discouraged overshoot, leading to "soft landings" averaging 12.5 minutes at induction (not favored by time-conscious OR schedules). Since induction and maintenance have competing goals in time and accuracy, a more effective solution might involve two independent, cooperative agents.

Conclusions

The RL agent demonstrated clinically suitable performance in the closed-loop control of propofol-induced hypnosis in healthy human volunteers. The agent achieved generalized control that compensated for varying degrees of intra- and inter-subject variation in propofol effect, suggesting that RL control may be applied to populations lacking good PK/PD models. Since the RL agent has demonstrated good control in both simulation and controlled human volunteer studies, the next objective is to assess agent performance in surgical patients.

Acknowledgments The clinical portion of this study was funded by the Department of Anesthesia, Stanford University School of Medicine. In addition, the authors would like to thank the operating room staff for their support, as well as Aspect Medical for providing an A-2000 BIS monitor.

References

Absalom, A., and Kenny, G. 2003. Closed-loop control of propofol anaesthesia using bispectral index(tm): performance assessment in patients receiving computer-controlled propofol and manually controlled remifentanil infusions for minor surgery. *British Journal of Anaesthesia* 90(6):737–41.

Absalom, A.; Sutcliffe, N.; and Kenny, G. 2002. Closed-loop control of anesthesia using bispectral index performance assessment in patients undergoing major orthopedic surgery under combined general and regional anesthesia. *Anesthesiology* 96(1):67–73.

Ausems, M.; Jr., C. H.; Stanski, D.; and Burm, A. 1986. Plasma concentrations of alfentanil required to supplement nitrous oxide anesthesia for general surgery. *Anesthesiology* 65(4):362–73.

Avidan, M.; Zhang, L.; Burnside, B.; Finkel, K.; Searleman, A.; Selvidge, J.; Saager, L.; Turner, M.; Rao, S.; Bottros, M.; Hantler, C.; Jacobsohn, E.; and Evers, A. 2008. Anesthesia awareness and the bispectral index. *New Eng J Med* 11(358):1097–1108.

Baird, L. 1995. Residual algorithms: reinforcement learning with function approximation. In *Proc. 12th International Conference on Machine Learning*, 30–37. Morgan Kaufmann.

Barvais, L.; Rausin, I.; Glen, J.; Hunter, S.; D'Hulster, D.; Cantraine, F.; and d'Hollander, A. 1996. Administration of propofol by target-controlled infusion in patients undergoing coronary artery surgery. *J Cardiothorac Vasc Anesth* 10(7):877–83.

Carregal, A.; Lorenzo, A.; Taboada, J.; and Barreiro, J. 2000. Intraoperative control of mean arterial pressure and heart rate with alfentanyl with fuzzy logic. *Rev Esp Anestesiol Reanim* 47(3):108–113.

Dayan, P. 1992. The convergence of TD(λ) for general λ. *Machine Learning* 8:341–362.

De Smet, T.; Struys, M.; Neckebroek, M.; den Hauwe, K. V.; Bonte, S.; and Mortier, E. 2008. The accuracy and clinical feasibility of a new bayesian-based closed-loop control system for propofol administration using the bispectral index as a controlled variable. *Anesthesia and Analgesia* 107:1200–1210.

Doufas, A.; Bakhshandeh, M.; Bjorksten, A.; Shafer, S.; and Sessler, D. 2004. Induction speed is not a determinant of propofol pharmacodynamics. *Anesthesiology* 101:1112–21.

Esmaeili, V.; Assareh, A.; Shamsollahi, M. B.; Moradi, M. H.; and Arefian, N. M. 2008. Estimating the depth of anesthesia using fuzzy soft computation applied to EEG features. *Intell. Data Anal.* 12(4):393–407.

Glass, P.; Bloom, M.; Kearse, L.; Rosow, C.; Sebel, P.; and Manberg, P. 1997. Bispectral analysis measures sedation and memory effects of propofol, midazolam, isoflurane, and alfentanil in healthy volunteers. *Anesthesiology* 86(4):836–847.

Gullapalli, V. 1993. Learning control under extreme uncertainty. In Hanson, S. J.; Cowan, J. D.; and Giles, C. L.,

eds., *Advances in Neural Information Processing Systems*, volume 5, 327–334. Morgan Kaufmann, San Mateo, CA.

Hu, C.; Lovejoy, W.; and Shafer, S. 1994. Comparison of some control strategies for three-compartment PK/PD models. *Journal of Pharmacokinetics and Biopharmaceutics* 22(6):525–550.

Leslie, K.; Absalom, A.; and Kenny, G. 2002. Closed loop control of sedation for colonoscopy using the bispectral index. *Anaesthesia* 57(7):690–709.

Lindholm, M.; Träff, S.; Granath, F.; Greenwald, S.; Ekbom, A.; Lennmarken, C.; and Sandin, R. 2009. Mortality within 2 years after surgery in relation to low intraoperative bispectral index values and preexisting malignant disease. *Anesthesia and Analgesia* 108(2):508–512.

Liu, N.; Chazot, T.; A, G.; A Landais, A Restoux, K. M.; PA Laloë, B. T.; Barvais, L.; and Fischler, M. 2006. Titration of propofol for anesthetic induction and maintenance guided by the bispectral index: closed-loop versus manual control: a prospective, randomized, multicenter study. *Anesthesiology* 104(4):686–695.

Moore, B.; Sinzinger, E.; Quasny, T.; and Pyeatt, L. 2004. Intelligent control of closed-loop sedation in simulated ICU patients. In *FLAIRS 2004*. AAAI Press.

Moore, B.; Pyeatt, L.; and Doufas, A. 2009. Fuzzy control for closed-loop, patient-specific hypnosis in intraoperative patients: A simulation study. In *Conf Proc IEEE Eng Med Biol Soc*, volume 1.

Myles, P.; Leslie, K.; McNeil, J.; Forbes, A.; and Chan, M. 2000. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. *The Lancet* 363(9423):1757–1763.

Pilge, S.; Zanner, R.; Schneider, G.; Blum, J.; and Kreuzer, M. 2006. Analysis of cerebral state, bispectral, and narcotrend indices. *Anesthesiology* 104(3):488–494.

Rampil, I. 1997. A primer for EEG signal processing in anesthesia. *Anesthesiology* 89(4):980–1002.

Röpcke, H.; Rehberg, B.; Koenen-Bergmann, M.; Bouillon, T.; Bruhn, J.; and Hoeft, A. 2001. Surgical stimulation shifts EEG concentration-response relationship of desflurane. *Anesthesiology* 94(3):255–113.

Sakai, T.; Matsuki, A.; White, P.; and Giesecke, A. 2000. Use of an EEG-bispectral closed-loop delivery system for administering propofol. *Acta Anesthesiologica Scandinavica* 44:1007–1010.

Sandin, R.; Enlund, G.; Samuelsson, P.; and Lennmarken, C. 2000. Awareness during anaesthesia : a prospective case study. *The Lancet* 355(9205):707–711.

Schaublin, J.; Derighetti, M.; Feigenwinter, P.; Petersen-Felix, S.; and Zbinden, A. M. 1996. Fuzzy logic control of mechanical ventilation during anaesthesia. *British Journal Anaesthesia* 77(5):636–41.

Schnider, T.; Minto, C.; Gambus, P.; Andresen, C.; Goodale, D.; Shafer, S.; and Youngs, E. 1998. The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers. *Anesthesiology* 88(5):1170–1182.

Sebel, P.; Bowdle, T.; Ghoneim, M.; Rampil, I.; Padilla, R.; Gan, T.; and Domino, K. 2004. The incidence of awareness during anesthesia: A multicenter United States study. *Anesthesia and Analgesia* 99:833–839.

Servin, F. 1998. TCI compared with manually controlled infusion of propofol: a multicentre study. *Anaesthesia* 53(Suppl 1):82–86.

Struys, M.; De Smet, T.; GreenWald, S.; Abasalom, A.; Bingé, S.; and Mortier, E. 2001. Closed-loop controlled administration of propofol using bispectral analysis. *Anesthesiology* 95(1):6–17.

Struys, M.; De Smet, T.; Greenwald, S.; Absalom, A.; Bingé, S.; and Mortier, E. 2004. Performance evaluation of two published closed-loop control systems using bispectral index monitoring: a simulation study. *Anesthesiology* 100(3):640–7.

Struys, M.; Coppens, M.; Neve, N. D.; Mortier, E.; Doufas, A.; Van Bocxlaer, J.; and Shafer, S. 2007. Influence of administration rate on propofol plasma-effect site equilibration. *Anesthesiology* 07(3):386–396.

Sutton, R., and Barto, A. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

Theil, D.; 3rd, T. S.; White, W.; Goodman, D.; Glass, P.; Bai, S.; Jacobs, J.; and Reves, J. 1993. Midazolam and fentanyl continuous infusion anesthesia for cardiac surgery: a comparison of computer-assisted versus manual infusion systems. *J Cardiothorac Vasc Anesth* 7(3):300–6.

Tsitsiklas, J., and Van Roy, B. 1996. An analysis of temporal difference learning with function approximation. Technical Report LIDS-P-2322, Massachusetts Institute of Technology.

Varvel, J.; Donoho, D.; and Shafer, S. 1992. Measuring the predictive performance of computer-controlled infusion pumps. *J Pharmacokinet Biopharm* 20:63–94.

Vuyk, J.; Engbers, F.; Burm, A.; Vletter, A.; and Bovill, J. 1995. Performance of computer-controlled infusion of propofol: an evaluation of five pharmacokinetic parameter sets. *Anesth Analg* 81(6):1275–1282.

Watkins, C. 1989. *Learning from Delayed Rewards*. PhD dissertation, Cambridge University, Computer Science Department.

Wood, W. 1989. Variability of human drug response. *Anesthesiology* 71(4):631–634.

Zadeh, L. 1965. Fuzzy sets. *Information and Control* 8:338–353.