Mechatronics 000 (2017) 1–9

Contents lists available at ScienceDirect

Mechatronics

journal homepage: www.elsevier.com/locate/mechatronics

Reinforcement learning-based thermal comfort control for vehicle cabins☆

James Brusey[a,*], Diana Hintea[a], Elena Gaura[a], Neil Beloe[a,b]

[a] Faculty of Engineering, Environment and Computing, Coventry University, Gulson Rd, Coventry, West Midlands CV1 2JH, United Kingdom
[b] Jaguar Land Rover Limited, Abbey Road, Whitley, Coventry, CV3 4LF, United Kingdom

ARTICLE INFO

Article history:
Received 2 January 2017
Accepted 10 April 2017
Available online xxx

Keywords:
Thermal comfort
Reinforcement learning
Equivalent temperature
Comfort model
Energy consumption

ABSTRACT

Vehicle climate control systems aim to keep passengers thermally comfortable. However, current systems control temperature rather than thermal comfort and tend to be energy hungry, which is of particular concern when considering electric vehicles. This paper poses energy-efficient vehicle comfort control as a Markov Decision Process, which is then solved numerically using Sarsa(λ) and an empirically validated, single-zone, 1D thermal model of the cabin. The resulting controller was tested in simulation using 200 randomly selected scenarios and found to exceed the performance of bang-bang, proportional, simple fuzzy logic, and commercial controllers with 23%, 43%, 40%, 56% increase, respectively. Compared to the next best performing controller, energy consumption is reduced by 13% while the proportion of time spent thermally comfortable is increased by 23%. These results indicate that this is a viable approach that promises to translate into substantial comfort and energy improvements in the car.

1. Introduction

Vehicle HVAC (Heating, ventilation, and air conditioning) systems aim to ensure that passengers are thermally comfortable. Traditionally, controllers for these systems are hand-coded and tuned to try to achieve this goal. However, there are a number of drivers for change:

1. Current systems only control cabin temperature whereas thermal comfort is also dependent on a multitude of other factors (such as radiant heat and airflow).
2. Past systems have relied on waste heat from the engine whereas electric vehicles produce much less heat and so a different design is required.
3. Current systems are energy hungry whereas electric and hybrid vehicles demand a much more energy efficient approach. Farrington and Rugh [12] report that air conditioning systems reduce the fuel economy of fuel-efficient cars by about 50%.

These drivers for change make redesign of many parts of the vehicle comfort delivery system timely. As this comfort system design changes, the controller must also adapt to best make use of the available actuation options.

The main idea in this paper is to show that Reinforcement Learning (RL) reliably produces a controller that uses less energy while delivering better comfort than existing hand-coded approaches (Section 4). We also show that the trade-off between energy and comfort can be adjusted to suit situations that demand either more comfort or better energy efficiency (Section 4.3.1). The approach requires a model of the cabin environment and we provide a simple, empirically validated, lumped model of the cabin's thermal environment (Section 3.1). The problem is then defined in terms of the state space (Section 3.3), action space (Section 3.5) and reward function (Section 3.6). Issues and implementation ramifications of this approach are discussed in Section 5.

2. Related work

2.1. HVAC control methods in vehicles

Much of the work on HVAC control [2,13,17,32] remains rooted in thermal comfort models developed for home and office indoor environments. The best known comfort model is the Predictive Mean Vote (PMV) [10,11,16], which estimates comfort based on: environmental parameters (such as air temperature, mean radiant temperature, relative air velocity and relative humidity); and personal parameters (such as metabolic rate and clothing thermal resistance). For example, Stephen et al. [32] derive a PMV-based fuzzy logic control mechanism, with rules like "if temperature is medium and activity is low, then PMV is near neutral".

☆ This paper was recommended for publication by Associate Editor Dr Reza Moheimani.

* Corresponding author.
E-mail address: j.brusey@coventry.ac.uk (J. Brusey).

http://dx.doi.org/10.1016/j.mechatronics.2017.04.010
0957-4158/

Although many aspects of vehicle thermal environment control are derivative of that in buildings, the vehicle's thermal environment is transient and non-uniform [37]. Thus it is recognised that what is appropriate in the thermal comfort model for a building may not be appropriate in a car [5,23].

While there are a number of thermal comfort models available, there is disagreement between these models about what contribution different parameters should have, or even what parameters to include [5]. Moreover, there are clearly parameters that might be considered but are not generally included. For example, occupants may enter the vehicle with latent or stored heat, they may have a physiological condition (such as a fever), or they may have cultural or personal preferences [4]. While there are many factors that can affect comfort, not all affect it equally. While air temperature remains central to comfort, as the number of sensors and intelligence of the controller within the car increases, it becomes possible to include more factors.

A number of additional *models, estimators*, and *predictors* populate the literature, typically accompanied by a strategy for HVAC control (e.g., Ueda and Taniguchi [36] predicts comfort based on facial skin temperature and cabin air temperature; Goenka and Maranville [17] proposed a zonal HVAC system driven on an occupant thermal comfort level based on sensor measurements, thermal comfort charts, the ASHRAE thermal scale, ISO 7730, the PMV index, the PPD index and their combination; Kranz [23] applies artificial intelligence methods to extract thermal comfort knowledge from the interaction between the passengers and the HVAC controls). Not surprisingly, most, if not all, of the proposed controllers are based on machine learning techniques. A prime reason is that car cabin comfort control is non-linear with respect to the observable state, for example: (a) the transfer of heat as a function of vent speed and vent temperature is non-linear; (b) any plant output limitation affects response in a non-linear fashion [8]; (c) comfort models, such as Predicted Mean Vote (PMV) and equivalent temperature (ET), are a non-linear function of their inputs.

Fuzzy logic is a common HVAC control approach given the imprecise nature of comfort [3,7,8,13,15,27,31,32,34] and many fuzzy-logic controllers have been found to perform better than the traditional air temperature controllers. Farzaneh and Tootoonchi [13] demonstrated that even better results were obtained when the parameters of the comfort oriented fuzzy controller were optimised by a genetic algorithm. Such controllers are, however, computationally expensive and can be difficult to design.

2.2. Reinforcement learning-based control applications

Dalamagkidis et al. [6] and Fazenda et al. [14] have examined the problem of optimising HVAC thermal comfort-based control through a RL-based technique in the context of buildings rather than cars. Dalamagkidis et al. [6] developed and simulated a reinforcement learning-based controller using Matlab/Simulink. The reward is a function of the building occupants' thermal comfort, the energy consumption and the indoor air quality. The proposed controller was compared to a Fuzzy-PD controller and a traditional on/off controller (an evaluation approach also applied here). The results showed that, after a couple of simulated years of training, the reinforcement learning-based controller performed better in comparison to the other two controllers.

Dalamagkidis et al. [6] highlight an issue with regard to reinforcement learning-based controllers—that of sufficient exploration. Taking random actions, even during short times, is unacceptable for a system deployed in a real environment and the authors recommend to exhaustively train the controller prior deployment and allow minimal or no exploration at all afterwards. This work provided inspiration and a good foundation for our work in vehicle cabins.

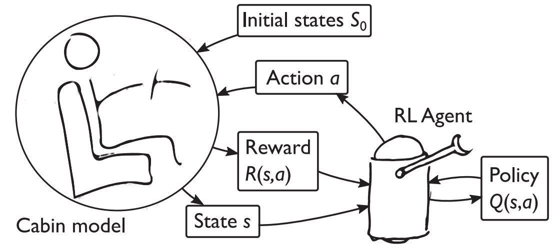

Fig. 1. The process of finding an optimal policy with RL involves modelling the cabin environment T, identifying the state S and action A spaces, defining the distribution S_0 of initial states, and defining an appropriate reward function $\mathcal{R}(s, a)$.

Fazenda et al. [14] have examined the problem of optimising comfort and energy using Q-learning with a state space that includes the time of day. They break the control problem down into: bang-bang control (when to turn the heater on or off) and set-point control (what temperature to request at what time). In their work, the tenant immediately responds to discomfort, which might seem unrealistic, but it provides similar input to the thermal comfort model used here. By including time, they neatly provide for pre-heating or cooling and this approach might also be used for the car cabin.

Less recently, Anderson et al. [1] have examined the problem of a simulated heating coil and combined a PI (proportional-integral) controller with an RL supervisor. They showed that the combined approach outperforms the base PI controller. This combination is similar to the approach here where the RL action is a vent temperature set-point that is passed to a base controller to achieve.

Dounis and Caraiscos [9] provide a detailed review of computational intelligence approaches in the built environment and show that, for the built environment, a variety of adaptive control approaches have been tried and advanced approaches (such as RL) have led to improved comfort and energy savings.

This past work demonstrates that RL, while untested, may be appropriate in this domain.

3. Materials and methods

We formulate the cabin comfort control problem (Fig. 1) as a Markov Decision Process (MDP) with continuous states defined by the tuple $\langle S, S_0, A, T, \mathcal{R}, \gamma\rangle$, where S is the (infinite) set of states of the cabin environment from which a set of initial states $S_0 \subseteq S$ is drawn, A is a finite set of actions (e.g., setting the blend door position), T: $S \times A \rightarrow S$ is a deterministic environmental model that maps states and actions to subsequent states, $\mathcal{R} : S \times A \rightarrow \Re$ is a function expressing the reward for taking an action in a particular state, and γ is a discount factor such that, for $\gamma < 1$, a reward achieved in the future is worth less than a reward achieved immediately.

The solution of the MDP is a policy π: $S \rightarrow A$ or mapping from states to actions and, in particular, an optimal solution is one that maximises the long-term, discounted expected reward. In algorithms such as Q-learning and Sarsa(λ), rather than find the policy directly, we estimate the expected value or utility $Q^\pi(s, a)$ of each state, action combination when following policy π. This expected value is the immediate reward $\mathcal{R}(s, a)$ plus the discounted subsequent reward, which can thus can be defined recursively,

$$Q^\pi(s, a) = \mathcal{R}(s, a) + \gamma Q^\pi(T(s, a), \pi(T(s, a))). \tag{1}$$

We can then progress greedily towards the optimum policy by updating the policy π to be that which maximises Q^π, or,

$$\pi(s) \leftarrow \arg\max_{a \in A} Q^\pi(s, a). \tag{2}$$

Since the policy for any state is easy to calculate from Q^π, it does not need to be explicitly stored.