A POMDP-based Multimodal Interaction System Using a Humanoid Robot

Sae Iijima
Advanced Sciences,
Graduate School of
Humanities and Sciences,
Ochanomizu University
g1220503@is.ocha.ac.jp

Ichiro Kobayashi
Advanced Sciences,
Graduate School of
Humanities and Sciences,
Ochanomizu University
koba@is.ocha.ac.jp

Abstract

In recent years, with the spread of the household robots, the necessity to enhance the communication capabilities of those robot to people has been increasing. The objective of this study is to build a framework for a dialogue system dealing with multimodal information that a robot observes. We have applied partially observable Markov Decision Process to modeling multimodal interaction between a human and a robot. Through the experiments, we have confirmed that our proposed framework functions properly and achieves effective multimodal interaction with a robot.

1 Introduction

In recent years, with the spread of the household robots, the necessity to enhance the communication capabilities of those robot to people has been increasing. Furthermore, we expect those robots which can observe information from multimodal resources and perform proper actions based on the observed information in interaction with people. In this context, the objective of our study is to achieve effective interaction with a robot using the multimodal information observed by the sensors of the robot. As a concrete system, we have implemented a dialogue system with the framework of partially observable Markov decision process (POMDP) in a humanoid robot called "Pepper" which can observe various multimodal information by its own sensors. Through several experiments, we aim to confirm that our system can assist Pepper to achieve flexible multimodal interaction with people.

2 Multimodal dialogue with a robot

2.1 Observation of multimodal information

In the experiments, we use a humanoid robot called "Pepper"[1] produced by SoftBank Co. Ltd. The figure of Pepper and its sensors are shown in Figure 1. We obtain multimodal information through the sensors of Pepper and aim to achieve multimodal interaction between Pepper and a user with those observed information. As for the multimodal information observed by the sensors equipped with Pepper, we obtain visual information from RGB camera, voice from microphone, contact information from touch sensor, and distance information from laser and sonar sensor. Pepper has face recognition function and can estimate user's age and identify five kinds of user's emotion: i.e., *neutral, happy, surprised, angry*, and *sad*, from user's facial expression.

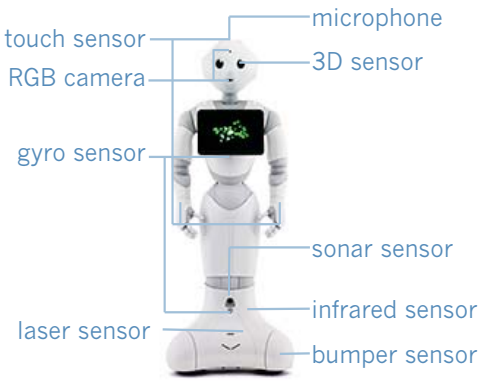


Figure 1: Pepper and its sensors

[1] http://www.softbank.jp/robot/

2.2 POMDP

In this paper, we use a framework of partially observable Markov decision process (POMDP) to represent uncertain states of the Markov decision process as stochastic states. The graphical model of POMDP is illustrated in Figure2.

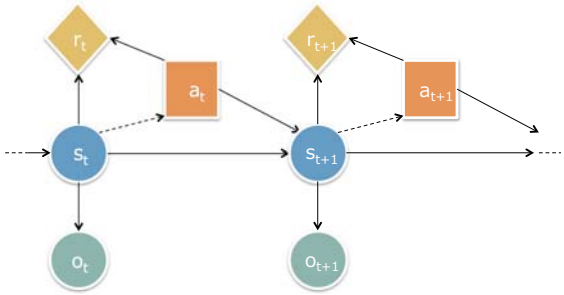


Figure 2: Graphical model of POMDP

A POMDP can be represented in the form of the following n-tuple:$\{S, A, T, O, Z, R, b_0\}$, where $s \in S$ denotes the state of a user, $a \in A$ is an action of the agent, $o \in O$ denotes an observation at state s. T is the probability of transitioning from state s to state s': $P(s'|s, a)$, Z is the probability of observing o' from state s' after taking action a: $P(o'|s', a)$, and $r(s, a) \in R$ is the reward signal received when executing action a in state s.

The process of POMDP is as follows: at each time-step, the target world is expressed as some unobserved state s. Because s is not known exactly, a distribution over states is maintained. This distribution is called "belief state" expressed as b, and its initial state is expressed as b_0. We represent $b(s)$ to indicate the probability of being in a particular state s. At each step, the belief state distribution b is updated as shown in equation (1).

$$b'(s') = k \times P(o'|s', a) \sum_s P(s'|s, a)b(s) \quad (1)$$

Here, k is regarded as a normalization constant to satisfy $\sum_s b'(s') = 1$.

2.3 Expansion to multimodal states

In the interaction between a user and an agent, we consider three states: s^e, s^p and s^l to represent user's emotional state, user's physical state, and user's intention by words, respectively. Here, o^e, o^p, o^l are the corresponding observations for those states, respectively. Figure3 shows the graphical model of the relation between states s and observations o.

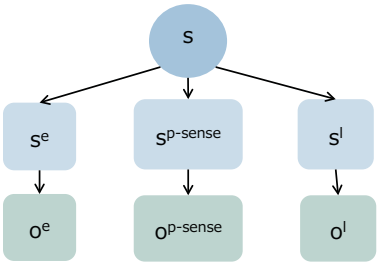


Figure 3: Graphical model of the relation between states and their observations

Here, a more detail about the multimodal states is explained as follows:

- Emotional state : s^e
  This factor indicates the state of user's emotion. We estimate this state based on the observation o^e observed by user's facial expression through an image recognition function equipped with Pepper.

- Physical state : $s^p(s^{p-dis}, s^{p-sense})$
  In our study, the physical state can be divided into two states. One is the distance state s^{p-dis} which represents the state of how far a user is from the agent, and the other is the state of sensing $s^{p-sense}$ which represents whether or not a user is touching the agent. We obtain observation o^{p-dis} from the laser sensor and the sonar sensor and $o^{p-sense}$ from the touch sensor equipped with Pepper.

- Linguistic state : s^l
  This factor indicates the state of user's intention provided by user's utterances. We obtain observation o^l through voice recognition system equipped with Pepper.

2.4 Stratified relation of states

In the case that it is difficult to obtain the optimal policy due to the increase of the state space in reinforcement learning, as one of the solution for this problem, the states are often reconstructed so as they are stratified (Dietterich, 2000). In the reinforcement learning employing stratified states in its de-

cision process, a complex task is divided into several subtasks which correspond to each strata of the stratified interaction processes. The agent learns the local policy at each strata and then learns the global policy for the complex task by unifying those local policies (Yamada et al., 2015).

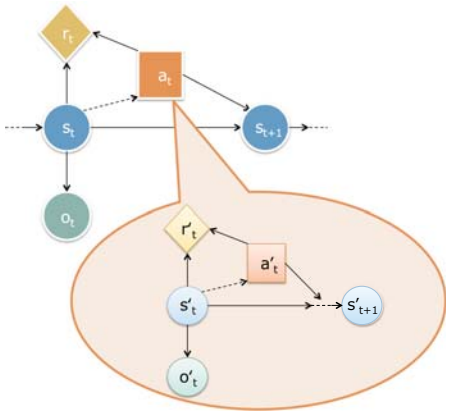


Figure 4: Stratified state in POMDP

2.5 Obtaining optimal policy by Q-learning

A plan to choose action a in state s is defined as policy π. Besides, π^* is defined as the optimal policy to choose optimal action a^* in state s. In POMDP, the states are represented in continuous states and therefore the number of states are monotonically increasing as the process unfolds. Therefore, Pineau et al. (2002) developed point-based value iteration algorithm to reduce calculation cost by transforming continuous values into discrete values at some points. In this paper, however, we assume that states s are regarded as being discrete for simplifying the model and then optimal policy π^* is obtained by Q-learning. The Q-values are updated as shown in equation (2). Here, α and γ indicate the learning rate and the discount rate, respectively.

$$Q(s,a) \leftarrow Q(s,a)+\alpha(r'+\gamma \max_{a'} Q(s',a')-Q(s,a)) \quad (2)$$

Figure 5 illustrates the introduction of Q-learning in the framework of POMDP to estimate the value of each state.

3 Experiments on multimodal iteraction

We conducted experiments employing a dialogue scenario in which Pepper and a user interact with

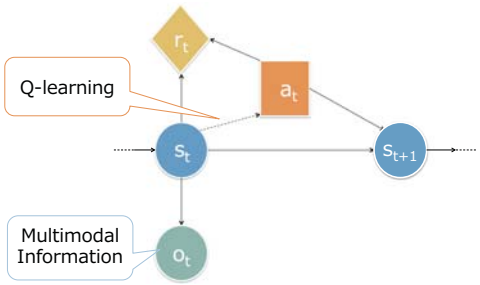


Figure 5: Q-learning in the framework of POMDP

multimodal information – the interaction is modeled by means of POMDP extended so as to be able to deal with multimodal information and to have stratified organization to represent the interaction. In the scenario, the task is stratified in accordance to the priority of interaction – here, user's location is the first priority to start interaction.

To build an interaction system, we improved the Python sample code[2] which implements a demo system for a spoken dialogue system with POMDP by (Williams et al., 2007) in the framework of PythonSDK[3] which is the software developer kit for Pepper.

3.1 The scenario of multimodal interaction

The scenario of multimodal interaction between Pepper and a user is shown in Table1.

Table 1: The scenario of multimodal interaction

Speaker	Interaction content	Observation	Action
User	(Far)	Distance	
Pepper	Come on, here!		Call
User	(Near)	Distance	
Pepper	Let's talk with me.		Speak to
User	Hello.	Voice	
Pepper	Hello.		Greet
User	(Sad face)	Picture	
Pepper	You look so sad! I will encourage you! (Pepper dancing)		Cheer
User	Thank you. (Patted head)	Voice Sensor	
Pepper	I am shy		Shy
User	(None)	Distance	
Pepper	(End of dialogue)		End

[2]https://github.com/mbforbes/py-pomdp

[3]http://doc.aldebaran.com/1-14/dev/python/index.html

3.2 Experimental settings

As for the interaction, in this study we represent the whole interaction in two stratified structure. The first strata represents the transition states of the physical location between Pepper and a user, and the second strata represents the dialogue interaction.

We show the detail settings of POMDP in the following – in this study, state transition probability, observed probability, reward, and the initial belief state are manually provided in advance.

- S : User's states

  S^{p-dis} : { None, Far, Near }
  S : { Greet, Sad, Fun, Happy, Unhappy }

- A : Actions

  A^{p-dis} : {None, End, Call, Speak to}
  A : { None, Greet, Cheer, Enjoy, Shy, Get down}

  Here, A^{p-dis} are the actions corresponding to S^{p-dis}.

- T : State transition probability, $P(s'|s,a)$
  The probabilities of transitioning from a state s to the next state s' for distance identification task and interaction task are shown in Table2 and 3, respectively.

Table 2: State transition probability for distance

s^{p-dis} / $s^{p-dis'}$	None	Far	Near
None	0.2	0.15	0.15
Far	0.2	0.15	0.15
Near	0.3	0.2	0.2

Table 3: State transition probability for dialogue

s / s'	Greet	Sad	Fun	Happy	Unhappy
Greet	0.2	0.25	0.25	0.15	0.15
Sad	0.2	0.15	0.15	0.25	0.25
Fun	0.2	0.15	0.15	0.25	0.25
Happy	0.3	0.2	0.2	0.15	0.15
Unhappy	0.3	0.2	0.2	0.15	0.15

- O: Observation information

  o^e : Observation of user's emotion from the facial expression through image recognition.
  o^{p-dis} : Observation of distance between Pepper and a user.
  $o^{p-sense}$: Observation of the sensing information of touch.
  o^l : Observation of user's voice.

- Z: Observation probability
  In accordance with the stratified relation of states, we consider two observation probabilities: $P(\boldsymbol{o'}|\boldsymbol{s'},a)$ and $P(o^{p-dis}|{s^{p-dis}}',a)$.
  In the experiments, we set the observation probability of user's voice as 0.8, and the other observation probability as 0.7.

- R : reward, $r(s,a)$
  The rewards given after every action for the identification of the distance to a user from Pepper and for the dialogue interaction are shown Figure4 and 5, respectively.

Table 4: R^{p-dis} : reward for the identification of distance

s^{p-dis} / a^{p-dis}	None	End	Call	Speak to
None	-1	5	-10	-10
Far	-1	-10	5	-10
Near	-1	-10	-10	5

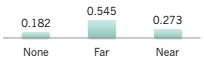
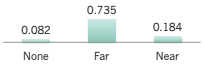
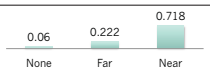
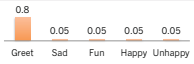
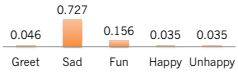
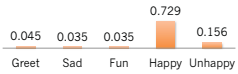
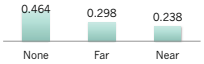
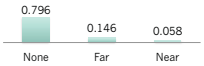
Table 5: R : reward

s / a	None	Greet	Cheer	Enjoy	Shy	Get down
Greet	-1	5	-10	-10	-10	-10
Sad	-1	-10	5	-10	-10	-10
Fun	-1	-10	-10	5	-10	-10
Happy	-1	-10	-10	-10	5	-10
Unhappy	-1	-10	-10	-10	-10	5

- ${b_0}^{p-dis}$ and b_0: The initial belief states
  ${b_0}^{p-dis}$ indicates the initial belief state of the distance between Pepper and a user, and b_0 indicates the initial belief state of a user.

  ${b_0}^{p-dis}$ = $(None : 0.2, Far : 0.2, Near : 0.2)$
  b_0 = $(Greet : 0.3, Sad : 0.2, Fun : 0.2, Happy : 0.15, Unhappy : 0.15)$

Table 6: Experimental result

Agent	Interaction Contents	Observation	$b(s)$	Action	reward
User Pepper	(Far) (None)	o^{p-dis}[Far]		None	-0.986
User Pepper	(Far) Come on, here!	o^{p-dis}[Far]		Call	1.100
User Pepper	(Near) Let's talk me.	o^{p-dis}[Near]		Speak to	0.819
User Pepper	Hello. Hello.	o^{l}[Greet]		Greet	2.541
User Pepper	(Sad face) How are you? I encourage you.	o^{e}[Sad]		Cheer	1.539
User Pepper	Thank you. (Patted head) I am shy	o^{l}[Happy] $o^{p-sense}$[Touch head]		Shy	1.965
User Pepper	(None) (None)	o^{p-dis}[None]		None	-0.989
User Pepper	(None) (End of dialogue)	o^{p-dis}[None]		None	1.964

- π^*: The optimal policy
  The optimal policy π^* shows the optimal action a^* in the belief state $b(s)$. It is represented in equation (3) by Q-value.

$$\pi^*(b(s)) = Q(b(s), a) \tag{3}$$

To find the optimal policy, we use ϵ-greedy method in Q-learning, and set the learning rate α as 0.2, and the discount rate γ as 0.9.

3.3 Experimental result

Table 6 shows an experimental result. Figure 6 shows the graphical model of POMDP for the scenario.

3.4 Discussions

Through the experiment, we have confirmed that our proposed multimodal interaction framework with a humanoid robot Pepper works well to interact with a user, and understood that the representation of the states in the interactive system tends to depend on the sensing functions and abilities of a robot. If each sensing function and ability is poor, it should be difficult to establish the interaction.

Furthermore, in this study we have built an interaction system with two strata in the framework of POMDP – we have set the first stratum so as it decides to start interaction based on the physical distance between Pepper and a user, and set the second stratum to control multimodal interaction. We have confirmed that the stratification of the whole interaction works well to reduce the dimension of states and then reduce calculation cost, and to make a good organization of the interaction.

In the experiment, we obtained the optimal strategy assuming that the states on the interaction are observed as being definite in order to reduce calculation cost. As a result, because the size of the scenario was short, there was not any problem in the interaction. However, we will have to take care of this problem, when we deal with a complicated and long interaction.

4 Related studies

As for employing POMDP in dialogue management, the essential features, e.g., how to model the inherent uncertainty in spoken dialog systems, why exact optimization is intractable, and how to describe the hidden information state model which does scale

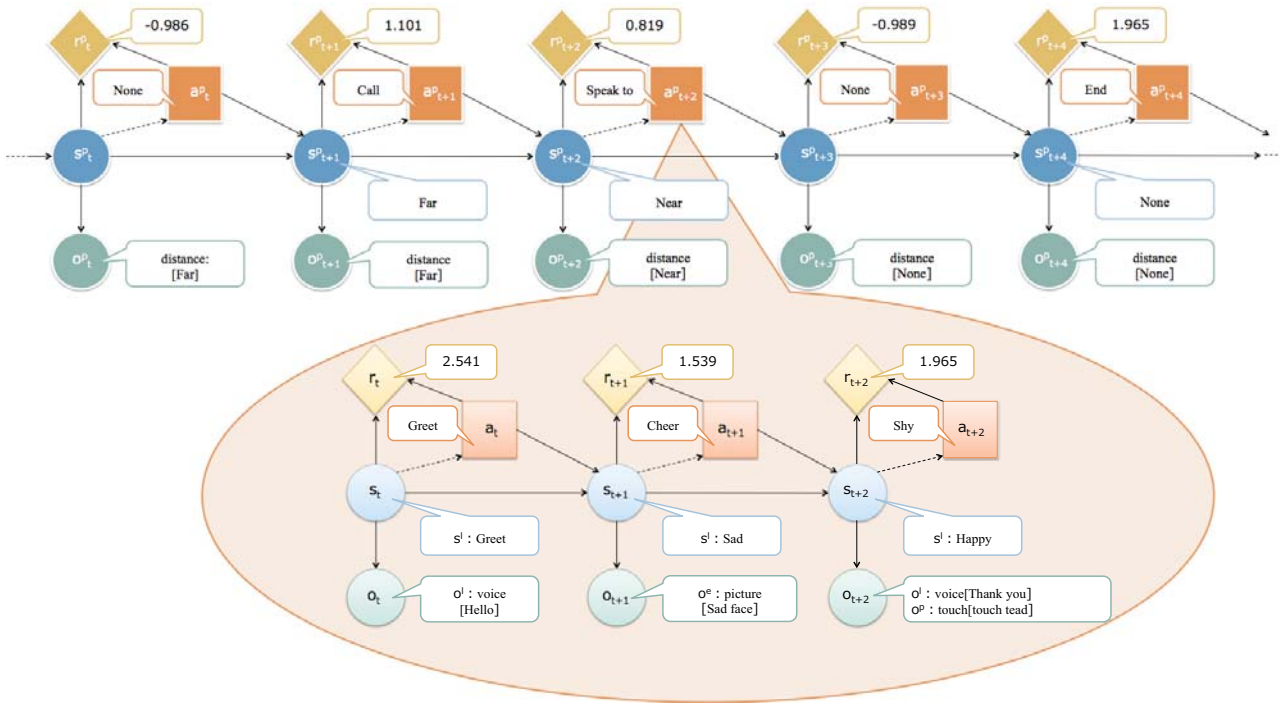


Figure 6: Overview of POMDP for the scenario

and which can be used to build practical systems, are studied in (Young, 2006; Young et al., 2007; Williams, 2006; Williams et al., 2007) – Young et al. (2007) partitioned the state space using a tree-based representation of user goals so that only a small set of partition beliefs needs to be updated at every turn to achieve the efficient calculation and showed a practical framework for POMDP-based spoken dialogue management system for the tourist information domain (Young et al., 2010).

Jurcicek (2011) proposed a reinforcement algorithm for learning parameters of dialogue systems modeled as POMDPs. Lison (2010) represented constraints on selecting actions with a small set of general rules expressed as a Markov Logic network in the framework of POMDP. He extended his idea to dialogue management based on the use of multiple, interconnected policies (Lison, 2010).

As for dealing with probabilistic states in a dialogue, a dialogue is modeled as Markov decision processes (MDPs) (Lemon, 2008; Lemon, 2011; Rieser, 2008; Rieser et al., 2009) and solved them by means of reinforcement learning (Sutton and Barto, 1998).

As a new trend in POMDP-based dialogue management, Gaussian Processes is applied to reinforcement learning Engel (2005) for optimal POMDP dialogue policies, in order to make the learning process faster and to obtain an estimate of the uncertainty of the approximation (Gasic et al., 2010; Gasic et al., 2013).

5 Conclusion

In this study, we have proposed a multimodal interaction system with a humanoid robot, expanding the framework of POMDP so as it can deal with multimodal information observed by the robot. In the system, we have achieved stratified interaction to reduce the increase of the user's belief states to deal with. Furthermore, we have also dealt with the estimated states as being definite so as to avoid the explosion of calculation cost. Through an experiment with a scenario, we have confirmed that our proposed method works well to achieve multimodal interaction between a user and a robot. As future work, we will consider the effective way to deal with

continuous states in the framework of POMDP employing multimodal information, and make a good organization of stratified structure in the dialogue interaction.

Acknowledgments

We would like to express our appreciation for financial support by Tateishi Science and Technology Foundation.

References

B. Bonet 2002. *An e-optimal grid-based algorithm for par- tially observable Markov decision processes* In Proc. of ICML, pp. 51-58.

T.G. Dietterich 2000. *An Overview of MAXQ Hierarchical Reinforcement Learning*, Lecture Notes in Computer Science, pp. 26-44.

Y Engel, S Mannor, and R Meir 2005. *Reinforcement learning with Gaussian processes*, In Proc. of the International Conference on Machine Learning.

M. Gasic, F. Jurcicek, S. Keizer, F. Mairesse, B. Thomson, K. Yu and S. Young 2010. *Gaussian Processes for Fast Policy Optimisation of POMDP-based Dialogue Managers*, In Proc. of SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 201-204.

M. Gasic, C. Breslin, M. Henderson, D. Kim, M. Szummer, B. Thomson, P. Tsiakoulis and S. Young 2013. *POMDP-based dialogue manager adaptation to extended domains* In Proc. of the SIGDIAL 2013 Conference, pp. 214-222.

F Jurcicek, B Thomson, and S Young 2011. *Natural actor and belief critic: Reinforcement algorithm for learning parameters of dialogue systems modelled as POMDPs*, ACM Transactions on Speech and Language Processing.

O. Lemon, 2008. *Adaptive natural language generation in dialogue using reinforcement learning*, In Proc. of the Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL), pp.141-148, London, UK. SemDial.

O. Lemon, 2011. *Learning what to say and how to say it:joint optimization of spoken dialogue management and natural language generation*, Computer Speech and Language, 25(2):pp.210-221.

P. Lison 2010. *Towards Relational POMDPs for Adaptive Dialogue Management* In Proc. of the ACL 2010 Student Research Workshop, pp. 7-12.

P. Lison 2011. *Multi-Policy Dialogue Management* In Proceedings of the SIGDIAL 2011 Conference, pp. 294-300

J. Pineau, G. Gordon, and S. Thrun 2002. *Point-based value iteration: An anytime algorithm for pomdps* In Proc. of the International Joint Conference on Artificial Intelligence, pp. 1025-1032.

V. Rieser and O. Lemon 2008. *Learning Effective Multimodal Dialogue Strategies from Wizard-of-Oz data: Bootstrapping and Evaluation*, In Proc. of ACL-08: HLT, pp. 638-646.

V. Rieser and O. Lemon, 2009. *Natural Language Generation as Planning under Uncertainty for Spoken Dialogue Systems*, In Proc. of the Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp.683-691, Athens, Greece.

N. Roy, J. Pineau and S. Thrun 2000. *Spoken Dialogue Management Using Probabilistic Reasoning*, In Proc. of the Association for Computational Linguistics.

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

J. D. Williams. 2006. *Partially Observable Markov Decision Processes for Spoken Dialogue Management*. Ph.D. thesis, University of Cambridge.

J. D. Williams, S. Young 2007. *Partially observable Markov decision processes for spoken dialog systems*, Computer Speech and Language, Volume 21 , Issue 2, pp. 393-422.

Y. Yamada, T. Takiguchi, and Y. Ariki, 2015. *SPOKEN DIALOGUE SYSTEM FOR PRODUCT RECOMMENDATION USING HIERARCHICAL POMDP*, 2015 First International Workshop on Machine Learning in Spoken Language Processing (MLSLP), 6 pages.

S. Young. 2006. *Using POMDPs for Dialog Management*. In Proc. of IEEE/ACL SLT, Palm Beach, Aruba.

J.Young, W. K. Schatzmann, and H. Ye. 2007. *The Hidden Information State Approach to Dialog Management*. In Proc. of 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07 (Volume:4), Honolulu, Hawaii.

S. Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu 2010. *The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management*, Computer Speech and Language, 24:pp. 150-174.