Decision Networks for Repair Strategies in Speech-Based Interaction with Mobile Tour-Guide Robots

Plamen Prodanov

Autonomous Systems Lab
Swiss Federal Institute of Technology, Lausanne
CH-1015, Lausanne, Switzerland
plamen.prodanov@epfl.ch

Andrzej Drygajlo

Signal Processing Institute
Swiss Federal Institute of Technology, Lausanne
CH-1015, Lausanne, Switzerland
andrzej.drygajlo@epfl.ch

Abstract – **The main task of a voice-enabled tour-guide robot in mass exhibition setting is to engage visitors in dialogue and provide as much exhibit information as possible in a limited time. In managing such a dialogue, extracting the user (visitor) goal or intention at each dialogue state is the key issue. In mass exhibition conditions uncooperative visitors and speech recognition limitations in noisy acoustic conditions may jeopardize user goal identification. In this paper, we introduce the use of sequential dialogue repair techniques, exploiting the inherent multimodality of the tour-guide robot, in order to reduce the risk of the resulting communication failures. Bayesian networks fusing acoustic and non-acoustic modalities during user goal identification serve as input to graphical models known as decision networks. Decision networks allow the definition of dialogue repair sequences as actions, and provide a decision-theoretic utility-based strategy for selecting actions. The benefits of the proposed repair strategies are demonstrated through experiments with the dialogue system of RoboX, a tour-guide robot successfully deployed at the Swiss National Exhibition (Expo.02).**

Index Terms – *Tour-guide robot, spoken dialogue, repair strategies, utility theory, decision networks.*

I. INTRODUCTION

The task of a mobile tour-guide robot in mass exhibition setting is to engage the visitors in an interactive tour, guiding them, moving autonomously, and presenting the items of the exhibition (exhibits). Such interaction is typically spontaneous and short-term, since visitors want to see as many exhibits as possible in a relatively short time [1], [2]. Therefore, the tour-guide robot should provide as much exhibit information as possible in limited time. To fulfil such a requirement human-robot interfaces should be intuitive, so that untrained and non-technical visitors to the exhibition can interact with the robot without prior instructions. Recent efforts were reported in this direction, investigating spoken dialogue with tour-guide robots in real exhibition conditions [2], [12]. The tour-guide robot equipped with speech recognition and synthesis system is able to sense changes in the acoustic aspect of its environment (e.g. speech) and perform speech-based actions, e.g. outputting synthesized speech. According to the exhibition plan and the time limitation, one interaction cycle usually aims at presenting a fixed number of exhibits, respecting the choice of the visitors [6], [2]. The tour-guide dialogue model is then constructed from a set of states, where the number of possible exhibit presentations per tour bounds the state space. At each state the task of the dialogue system is to infer the goal of the visitor through the speech recognition system, i.e. his/her intention to attend the

possible next state presentations in order to decide on which exhibit to present next.

The operating conditions in a mass exhibition environment abound with a variety of uncertainties [1], [15]. First, visitors are not always acting cooperatively during the tour-guide dialogue. For example, they might miss the right moment to answer to the robot or simply leave before the end of interaction. There are even cases when they try to confuse the tour-guide robot for fun [2], [15]. Such behaviours make visitors' intentions difficult to anticipate in human-robot interaction, causing ambiguity and errors when the robot has to interpret them. Second, data coming from the robot's input modalities and in particular the speech signal captured by the microphone, can be very noisy. The presence of a crowd of people and moving robots in the exhibition room results in adverse acoustic conditions, causing errors in the speech recognition. Communication failures may arise in dialogue due to the above two factors. Hence a system managing speech-based interaction with visitors should employ error-handling techniques for reducing the effect of unreliable speech recognition in dialogue, as well as the effect of user goals leading to communication failures.

A. Error handling in speech-based human robot interaction

Standard techniques for error handling in speech recognition are based on detecting errors using the recognition scores and correcting them through repair dialogues [10], [14]. Detecting errors using only speech recognition can be difficult and repair dialogues may be inefficient in the acoustic conditions of mass exhibition. In our previous work [11], [12], we have outlined the advantages of statistical modality fusion for correcting speech recognition errors in the process of identifying user goals in the tour-guide dialogue. The chance for communication failure can be explicitly modelled including an "undefined user goal" in the set of possible user goals. Bayesian networks can be used to elicit a probability for the undefined user goal at each dialogue state, fusing acoustic (speech recognition result (score)) as well as non-acoustic aspects (features derived from SICK laser scanner signal) of the user goal [12]. Given the probability distribution over the possible user goals and in particular the undefined user goal, the error handling strategy should decide explicitly whether to apply a dialogue repair sequence. Additionally in the case of tour-guide robots such sequence may employ speech-based as well as other modalities, e.g. move event [5], [2]. For example an evidence of presence or absence of communicating visitors, given by the laser scanner signal

can result in a repair sequence, including an initial phase in which the robot moves around searching for a visitor.

If the dialogue repair sequences are defined as actions that the robot can perform, some principles from decision theory provide explicit way of selecting actions, given the robot's preferences and the level of uncertainty in user goal identification at each dialogue state. Decision theory defines action selection strategies based on explicit measure of robot's action preferences named utilities and the principle of maximum expected utility (MEU) [10], [13]. This principle can be implemented using the concept of the "decision networks" - a graphical representation that combines Bayesian networks with additional node types for actions and utilities. In this paper, we report on the use of decision networks for modelling repair strategies fitted to the requirements of speech-based interaction with a tour-guide robot.

The paper is structured as follows: Section 2 provides the necessary background information on maximum expected utility (MEU) and decision networks. Section 3 and 4 describe how the MEU principle can be used in the context of a real tour-guiding application, using the formalism of decision networks. Finally the potential benefits of the decision theoretic strategies for tour-guiding dialogue repairs are outlined with future perspectives in the Discussion and Conclusion sections of the paper.

II. DECISION THEORY

A. Maximum expected utility

The principle of maximum expected utility (MEU) is used in decision theory and modern artificial intelligence for modelling the action selection strategy of utility-driven agents [13].
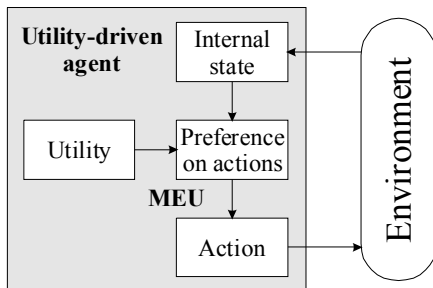

Fig. 1 Architecture for a utility-driven agent.

Fig. 1 depicts the architecture for a utility-driven agent. Such an agent maintains an internal state (model) of its environment given its sensors' information. On the other side a utility function is used to model the agent's preferences for the different actions through which the agent can manipulate the environment. The utility function assigns a numerical value to each agent's actions, given the state of the environment. Finally, the process of action selection is modelled by combining probability and utility theory. Probability theory is used to model the agent's internal state, given the information (evidence) extracted from its sensors. Utility theory is used to model the agent's preferences between the states of the external environment, or in other words the part of the world perceived by its sensors and modelled by the agent's internal state. These preferences are captured by the utility function mentioned above. This

function assigns a number to rank the agent preference for a given state, which can result from a given executed action. We will use $U(s, a)$ to denote the utility of a state s given that the agent has performed an action a. $P(S=s|E)$ will denote the probability of each state value, given the current evidence E from the sensor data. Then the maximum expected utility is given by the following equation:

$$MEU(A \mid E) = \arg\max_a \sum_s P(S = s \mid E) \cdot U(s, a). \qquad (1)$$

The maximum expected utility principle in decision theory states that an intelligent agent should choose an action that maximises the expected utility of that action, given the sensor evidence for the state of the world at the instance of decision-making. This kind of utility driven decisions can be implemented and visually represented with the help of decision networks [10], [13], also known as influence diagrams.

B. Decision networks

In a decision network (*DC*) there are three types of nodes, i. e. chance nodes (ovals), decision nodes (rectangles) and utility nodes (diamonds). An example of a decision network is shown in Fig. 6. The chance nodes represent random variables. The agent is usually uncertain about the exact values of these variables. Some of the chance nodes can represent features extracted from the agent sensors; others can represent different aspects of the agent internal state. Decision nodes represent possible choice of actions. The utility nodes represent the utility function. Since the utility function depends on the agent's internal state and the actions, utility nodes usually have one or more chance nodes and the decision node as parents. Bayesian networks [7], [13] are often used to model the probabilistic dependences between the chance nodes and serve as an input to the decision network. They can be used to produce probability values on the state variables, e.g. $P(S=s|E)$. Then applying (1) will result in selecting the action with MEU, given the set of possible actions.

To construct a decision network for a particular decision problem a precise definition of the agent's internal state, actions and preferences (utilities) is derived from the requirements of the agent's task. In order to derive these requirements and define the above terms in the case when the agent is a tour-guide robot some specific details concerning the robot platform and the speech-based human-robot interaction in mass exhibition conditions are needed.

III. TOUR-GUIDE DIALOGUE

We take as an example the interactive tour-guide robot RoboX (Fig. 2) successfully deployed at the Swiss National Exhibition (Expo.02) [2], [6]. For the purpose of human-robot interaction RoboX is equipped with the following modalities: speech recognition system, interactive buttons, video camera as input modalities, and led matrix animations, expressive face, speech synthesis system as output modalities. For the purposes of navigation and obstacle avoidance the robot is additionally equipped with two laser scanners (laser range finders SICK), emergency stop button, and bumpers for avoiding collision with obstacles that cannot be detected by the laser scanner beam [5]. During Expo.02 eleven robots interacted with individual visitors as

well as crowds of people (hundreds of thousands of visitors during 5 months, seven days a week, 10 hours per day).
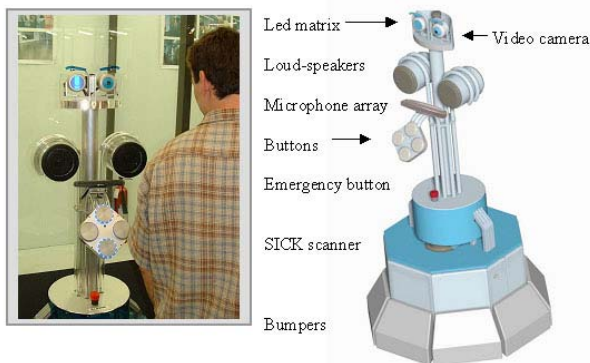

Fig. 2 The mobile tour-guide robot RoboX.

People were not instructed beforehand as how to operate the robot. In such conditions it is preferable that the tour-guide robot take the initiative in the dialogue [2]. Thus, a successful tour-guide robot should be capable of detecting the presence of people, engaging them in dialogue, presenting the items of the exhibition (exhibits). During this interaction the visitors' behaviour can vary from collaborative to investigative and even "destructive" [2]. The tour-guide robot needs to interpret this behaviour into "user goals" relevant to the tour-guide dialogue. The tour-guide dialogue can be represented as a set of dialogue states, where each dialogue state corresponds to a sequence of low-level behavioural events, such as a speech synthesis event, a speech recognition event, a robot movement event, etc. The sequence of events forming a dialogue state is organized to present a specific exhibit. Thus the number of dialogue states is fixed and can be defined in advance based on the number of exhibits described by the particular exhibition plan. Each dialogue state contains verbal interaction in the form of question/response pair, during which the speech recognition is typically used to infer the "goal" of the speaker in the context of the current state [4]. We assume that the spoken utterances coming from visitors during interaction can be mapped into a finite number of state dependent user goals, which are used to infer the next dialogue state. The question/response pair in the case of RoboX is at the beginning of each exhibit's presentation and consists of yes/no question from the robot and answer from visitor; e.g. the tour-guide robot asks the visitors if they want to see the next exhibit. One complete tour is limited to five question/response pairs (user goal/actions decision points) [6] in order to provide visitors with approximately 15 minutes interaction time. The time limitation factor is motivated by the short-term nature of the human-robot interaction in exhibition conditions. The exact value resulted from requirements concerning the visitor flow management during Expo.02. Given the main task of the user guide, e.g. presenting as much exhibit information as possible, the average number of exhibit presentations, resulting from correctly recognized responses, can be used as a measure for successful interaction.

The initial experiments during Expo.02 showed that such an interaction scheme could be seriously challenged by the visitors' behaviour. It is often the case that people do not follow the choice suggested by the robot, using out-of-vocabulary words and even giving both *yes* and *no* answers or simply remain silent [2]. Therefore the speech recognition system of RoboX needs to distinguish between the keywords *yes*, *no* and out-of-vocabulary words, fillers, coughs, laughs and general acoustic phenomena different from the keywords, called garbage words (*GB*). The *O*bserved *R*ecognition *R*esult *ORR*={*yes*, *no*, *GB*} is then mapped into three possible user goals (*UG*), accounting for the visitor intention: "the user is willing to see the next exhibit" (*ORR*=yes then *UG*=1); "the visitor is unwilling to see the next exhibit" (*ORR*=no then *UG*=2) and "user goal is undefined" (*ORR*=GB then *UG*=0) (Fig. 3).
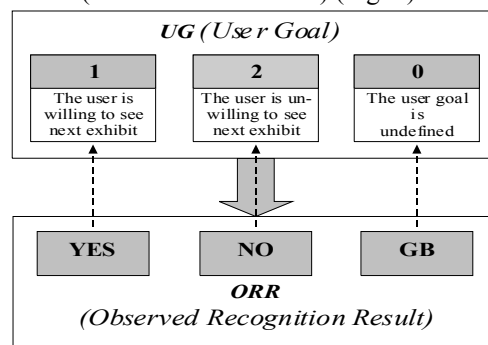

Fig. 3 Mapping of observed recognition result and user goal values.

During Expo.02 the background acoustic noise and unexpected visitor behaviour caused significant errors in recognizing the *GB* word. An example case was when initially interested visitors were leaving the robot to respond to other people calling them. When this unexpected behaviour was coinciding with the question/response pair, the *GB* word was often misrecognized for *yes* or *no* answer by the robot. In order to infer the right user goal (*UG*=0) in this case, auxiliary information from the laser scanner signal revealing presence of visitors in close distance with respect to the robot's microphone array (<1.5m) proved to be beneficial.

A. Multimodal user goal identification with Bayesian networks

The speech recognition result can be seen as the "acoustics-related" aspect of the user goal. Another "spatial" aspect of the user goal is provided by the laser scanner signal, which contains information about the location of a visitor communicating with the robot. This "spatial" aspect of the user goal is essential, as absence of a user in given range in front of the robot could signal possible communication failure. To incorporate the spatial aspect of the user goal, we define the binary event *U* "user in range for communication" (*U*=1 user is in range, *U*=0 user is out of range). Combining the observed recognition result (*ORR*) with evidence from the acoustic noise insensitive *L*aser *S*canner *S*ignal (*LSS*) that can affect the event *U*=1 can change the "confidence" about the result of speech recognition. On the other hand, to define the influence of the acoustic environment on the speech recognition reliability we define the binary event *DR* "data reliability" (*DR*=1 acoustic data is reliable, *DR*=0 acoustic data is unreliable). To infer the state of *DR* the tour guide robot needs additional evidence about changes in the environment that can affect the reliability of the incoming

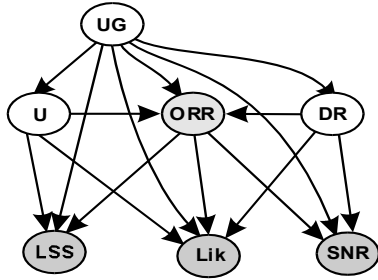data and in particular the effect of acoustic noise on the speech signal.

Fig. 4 Bayesian network for multimodal user goal identification

The likelihood (*Lik*) of the observed recognition result along with an estimate of the speech-to-noise ratio (*SNR*) of the captured acoustic signal can provide information about the environmental acoustic conditions [3]. The *U* and *DR* events can directly influence *ORR*. Since these influence are not deterministic the causal relationships should be seen as probabilistic. It can happen that people are near the robot, speaking as expected and recognition errors still occur. Therefore Bayesian networks can be used to fuse data coming from the speech recognition and the laser scanner modalities of RoboX for inferring the user goals of the visitors in a multimodal fashion.

A Bayesian Network (*BN*) [7], [13] is a graphical model used to describe dependencies in a multivariate probability distribution function (pdf) defined over a set of random variables. The topology of the network is defined by a *D*irected *A*cyclic *G*raph (*DAG*), consisting of nodes corresponding to the variables and arcs representing the conditional dependence assumptions between the variables. The arcs point in the direction from the cause to the consequence or from the parent variable to its children.

Fig. 4 depicts the Bayesian network used for the multimodal user goal identification, where shaded nodes mark observed (evidential) variables. The design methodology for constructing this network is presented in detail in [11] and [12]. For the goal of this paper it suffices to note that if we define the topology and the conditional probability distribution functions for all nodes given their parents, inference on each node in the network can be done [8], [9]. The task of inference is to compute a posterior distribution for a set of "query" variables, given some set of observed variables, e.g. $P(UG|E)$, $P(DR|E)$, $P(U|E)$ etc., where $E = \{LSS, Lik, SNR, ORR\}$ for the *BN* in Fig. 4. These posterior distributions can be used like chance nodes in a decision network, making a Bayesian network a suitable input for a MEU-based decision system.

IV. Decision Networks for Tour-Guide Repair Strategies

In the context of utility driven tour-guide robot the user goal values can represent the agent's (robot) internal states. The tour-guide dialogue can be seen then as a process of decision-making, where each state in dialogue is considered as a decision point. At each decision point the "question/response" pair is used to probe the external environment and elicit a probability distribution over the robot's internal states - $P(S|E)$. In the previous section we have equipped the tour-guide robot with a model for its

internal state, e.g. a Bayesian network for estimating $P(S|E)$, where $S=UG$, and $E =\{LSS, Lik, SNR, ORR\}$. In order to apply (1), we still need to define precisely the set of robot's actions and the utility function.

A. Defining actions and repair strategies

The dialogue sequences presenting the exhibits in one complete tour can be seen as valid dialogue actions for the case of *UG*=1. We will refer to these sequences as "Present next exhibit" actions. On the other hand, the question/response pairs offering exhibit presentations to the visitors can be seen as valid actions for the case of *UG*=2. We will refer to these actions as "Offer another exhibit" actions.

It has been outlined in Section I and III that due to uncooperative visitors and adverse acoustic conditions during dialogue, the visitor's intentions cannot always be classified into meaningful user goals in the context of tour guiding (e.g. simple accept/reject responses in the case of RoboX). In this case using an "undefined" user goal (*UG*=0) is well motivated and requires "repair" actions that the robot can perform to avoid communication failures. To define these "repair" actions, we take into account the tour-guiding dialogue requirements as outlined in Section III: provide exhibit information through efficient speech-based interaction in limited time, where the number of presented exhibits, after correct user goal identification, can be used as a measure for efficient interaction.
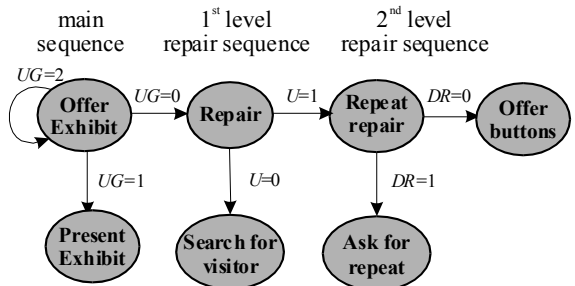

Fig. 5 Tour-guide dialogue state transition diagram.

Dialogue repair sequences generally occur as a side (unexpected) sequence in the normal process of human-robot interaction and may lead to delays in communication. Therefore, given the above task requirements the "repair" actions should avoid unnecessary repetitive patterns that might often arise using speech recognition in noisy acoustic conditions. In building "time-saving" repair sequences using alternative input and output robot modalities can be very beneficial. For example, in the case of an absence of communicating visitor (*U*=0) the most appropriate repair sequence should include an initial phase in which the robot moves around searching for a visitor. We will define such a repair sequence as the "Search for visitor" action. In the case of *U*=1, performing a "Repeat repair" action, e.g. asking the user for repeated input trial would be the fastest possible repair sequence. However knowing that *U*=1 and *DR*=0 would give less motivation to the use of a speech-based "Ask for repeat" repair action, compared with an alternative use of the interactive buttons through the "Offer buttons" repair action. Fig. 5 depicts the state transition diagram of the tour-guide dialogue, including a two-level repair strategy as outlined in the above recommendations. In

real conditions however the states of *UG*, *U* and *DR* are never known for sure. If *UG*, *U* and *DR* are seen as chance nodes, decision networks can be used as a state transition model for selecting valid actions using the principle of maximum expected utility (MEU), given by equation (1).

B. Decision networks for tour-guide repair strategies

Fig. 6, 7 and 8 depict the decision networks *DC*1, *DC*2 and *DC*3 that can be used for selecting actions in the three decision levels of the tour-guide dialogue in Fig. 5. The same Bayesian network (Fig. 4) is used as an input for the three decision networks to output values for the corresponding posterior distributions needed for equation (1), e.g. $P(S|E)=P(UG|E)$ in the main dialogue sequence case (*DC*1), $P(S|E)=P(U|E)$ for the 1st level (*DC*2), and $P(S|E)=P(DR|E)$ for the 2nd level (*DC*3) of dialogue repair, given the evidence $E =\{LSS, Lik, SNR, ORR\}$. Finally, the utility functions associated with the utility nodes in the three networks that are used in calculating the corresponding actions' expected utilities are defined as real valued tables, indexed by the actions and user goals.
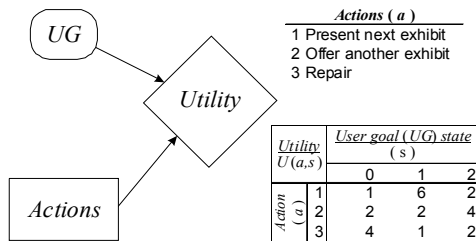

Actions (*a*)
1 Present next exhibit
2 Offer another exhibit
3 Repair

Utility $U(a,s)$	User goal (UG) state (s)		
	0	1	2
Action (a) 1	1	6	2
2	2	2	4
3	4	1	2

Fig. 6 *DC*1: decision network for managing the main tour-guide dialogue sequence

Actions
1 Repeat repair
2 Search visitors

Utility $U(a,s)$	U state (s)	
	0	1
Action (a) 1	1	5
2	5	3
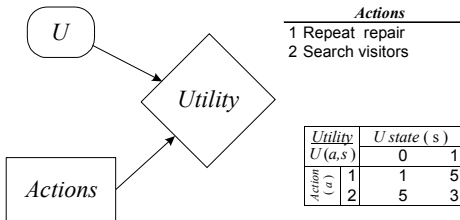
Fig. 7 *DC*2: decision network for managing the first repair level.

Actions
1 Ask for repeat
2 Offer buttons

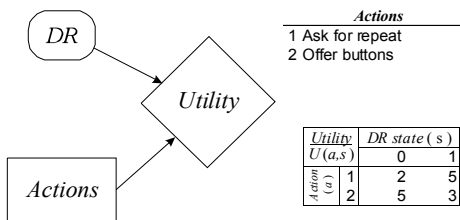| Utility $U(a,s)$	DR state (s)	
	0	1
Action (a) 1	2	5
2	5	3

Fig. 8 *DC*3: decision network for managing the second repair level.

The numerical values of utilities in general are mathematically unique up to a positive affine transformation such that if $U(x)$ is the utility, then $k_1U(x)+k_2$ is equivalent for any constant $k_1>0$ and k_2 [10]. The particular $U(s, a)$ values in the utility tables for above three decision networks represent the tour-guide preferences over its actions, given the user goal values and are motivated by the tour-guiding requirements outlined in IV.A. These values can be interpreted as rewards that the tour-guide robot would gain in performing particular action, given the user goal values at the current decision point. For example, due to the time

limit during interaction the most preferable action for a "rational" tour-guide robot would be to "Present next exhibit" in the case of *UG*=1, and the least preferable one would be the "Repair" action, since it might lead to unjustifiable delays in interaction. However, in the case of *UG*=0 performing the "Repair" action would be much more relevant in order to prevent communication failure. Given the utility tables, formula (1) can be used by the three decision networks in the order specified in Fig. 5 to select the actions that maximize the expected utility of that action, given the distribution over the values of the corresponding chance nodes {*UG*, *U*, *DR*}.

C. Experiment with data from Expo.02

During Expo.02 we collected multimodal data samples from the interactive tours of RoboX with the visitors (audio recordings and laser scanner readings) [11]. The data were manually labelled with *UG*={0, 1, 2} (Section III.A) by human experts. Approximately 50% of these samples were labelled with *UG*=0. We have trained the *BN* in Fig. 4 on a portion of 810 examples that resulted after balancing uniformly the *UG* values. Another balanced portion of 390 samples was used for testing the *BN* and results were reported in [11]. In order to test the benefits of the proposed repair strategies, we have performed tests with only the data of *UG*=0. We have used 130 testing examples (Fig. 9) containing values of the three posteriors $P(UG|E)$, $P(U|E)$ and $P(DR|E)$ calculated by the BN in Fig. 4 for 130 cases of an undefined user goal (*UG*=0). The decision network *DC*1 was used initially to decide if a repair action is needed. In the case, when the repair action had maximum expected utility, *DC*2 was used to decide if there is a visitor in front of the robot and consequently *DC*3 in order to decide what input modality to be offered to the user during the repair sequence. The results from the experiment are shown in Table 1.

V. DISCUSSION

As can be seen from Table 1 in 83% of the cases the network *DC*1 has correctly assigned a repair action, and 94% of the repair actions correspond to "Search visitors" actions. At the end, in all the 6 cases in which the user was estimated to be present, he/she is reoffered to use the speech modality during the final repair action. Given that visitors might utter out-of vocabulary words at that point the "Ask for repeat" action may lead to delays in conversation. To handle this issue making the utilities dependent on the number of times an action is executed (e.g. $U_t < U_{t-1}$) might be beneficial [10]. Finally, in 102 out of 130 cases the mobility of the tour-guide robot provides an efficient way to avoid sure communication failure due to the absence of visitor during interaction. We also see that among the wrongly selected actions at the main dialogue sequence level, the decision network *DC*1 has selected in most of the cases the "Offer another exhibit" action. This decision can be seen as safer, compared with the "Present next exhibit " action when there is no audience in front of the robot. This safer action selection strategy is explicitly encoded through the corresponding utility values in *DC*1 (Fig. 6). Such decision theoretic repair strategies provide substantial degrees of freedom in modelling the tour-guide behaviour.
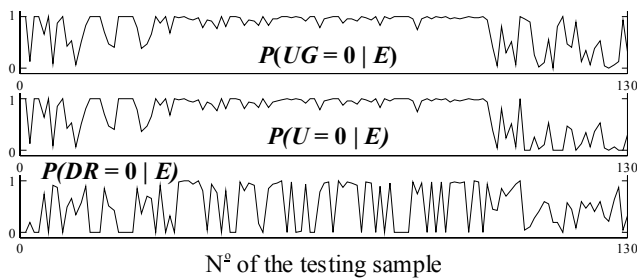
Fig. 9 Graphical representation of the chance nodes' probabilities in $DC1$, $DC2$ and $DC3$ for 130 examples of $UG=0$.

TABLE 1
EXPERIMENTAL RESULTS

Main sequence (DC1)		MEU action	Total
Actions:	1 Present next exhibit	9	6.9%
	2 Offer new exhibit	13	10.0%
	3 Repair	108	83.1%
1st level repair (DC2)		MEU action	
Actions:	1 Repeat repair	6	5.6%
	2 Search visitors	102	94.4%
2nd level repair (DC3)		MEU action	
Actions:	1 Ask for repeat	6	100%
	2 Offer buttons	0	0%

Given equally likely user goals the MEU principle will select the action with the maximal sum of the utilities across all user goals (the sum of the rows in the utility tables). In that sense the individual $U(a, s)$ values also contribute to the global importance (preference) on actions. Following such global preferences the behaviour of the tour-guide robot during interaction can be adapted to be more conservative or less conservative in performing the repair actions. For example in Fig. 6 the global importance of presenting exhibits is higher compared to the one of offering a new exhibit or the repair option. Since searching for visitors might encourage the visitors around the robot to join the interaction, the global preference is in the favour of the "Search visitors" action in the first level of the tour-guide repair strategy (Fig. 7). In the decision network corresponding to the second repair level (Fig. 8), i.e. "Ask for repeat" vs. "Offer buttons" the second action can be seen as globally more preferable. Since button's input during speech-based interaction does not depend on the acoustic noise, it is considered as more reliable at high levels of acoustic noise.

User satisfaction tests with the basic (yes/no) recognition system during Expo.02 were presented in [2]. Since RoboX is at present in a phase of reprogramming, we will provide such tests for the decision networks based dialogue system in our future work.

VI. CONCLUSION

In this paper, we presented a complete methodological concept for designing and implementing of repair strategies for avoiding communication failures in spoken dialogues with mobile tour-guide robots in mass exhibition conditions. In these conditions non-collaborative visitors' behaviour and adverse acoustic conditions were shown to be among the main factors for communication failures in speech-based interaction. The problem of tour-guide dialogue management is shown to depend on a robust inference of the user goal at each dialogue state, where the chance for communication failure can be explicitly modelled through an "undefined user goal". Bayesian networks are used to elicit probability distribution over the set of user goals, fusing acoustic (speech recognition result) and spatial (laser scanner signal) aspects of the user goal. In the case of high probability for the undefined user goals, dialogue repair sequences were chosen in accordance with the tour-guide requirements, exploiting the potential benefit of different input and output robot modalities, e.g. speech or buttons-based input, move event, etc.

Given that the real state of the user goal is never known for sure by the robot, the strategies for repair-action selection can be modelled using concepts from probability and decision theory and related visual representations, e.g. Bayesian networks and their extensions - decision networks. Decision theory allowed us to define the tour-guide dialogue as a sequential process of decision-making, where decision networks were used to choose from the available actions at each dialogue state (decision point). Decision networks utilise a mathematical framework for choosing actions, based on the maximum expected utility (MEU) of the repair actions over the distribution of the user goals given by the Bayesian network. The MEU principle allows modelling of complex task-oriented tour-guide behaviours, through manipulating the utility function values. In the paper it is shown that decision networks can be used for modelling a variety of tour-guide repair strategies, taking into account different aspects of the user goal.

REFERENCES

[1] W. Burgard, et al., "Experiences with an interactive museum tour-guide robot," *Artificial Intelligence*, vol. 114 (1-2), pp. 1-53, 1999.
[2] A. Drygajlo, et al., "On developing voice enabled interface for interact-tive tour-guide robots," *Advanced Robotics*, vol. 17 (7), pp. 599-616, 2003.
[3] X. Huang, Al. Acero, Hsiao-W. Hon, *Spoken Language Processing*, Prentice Hall PTR, 2001.
[4] Er. Horvitz, T. Paek, "A computational architecture for conversation," *Proc. of the 7th Int. Conf. on User Modeling*, Banff, Canada, June 1999, pp. 201-210.
[5] B. Jensen, et al., "The interactive autonomous mobile system RoboX", *Int. Conf. on Intelligent Robots and Systems*, IROS 2002, Lausanne, Switzerland, Sept. – Oct., 2002, pp. 1221-1227.
[6] B. Jensen, et al., "Visitor flow management using human-robot interaction at Expo.02," *Workshop: Robotics in Exhibitions*, IROS 2002, Lausanne, Switzerland, Oct. 2002.
[7] F. Jensen, *An Introduction to Bayesian Networks,* UCL Press, 1996.
[8] F. Jensen, J. D. Lafferty, R. L. Mercer, "Bayesian updating in causal probabilistic networks by local computations," *Computational Statistics Quarterly*, vol. 4, pp. 269-289, 1990.
[9] K. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. thesis, U. C. Berkeley, July 2002.
[10] T. Paek, Er. Horvitz, "On the utility of decision-theoretic hidden subdialogues," *Proc of ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, Chateu-d'Oex-Vaud, Switzerland August 28-31, 2003.
[11] Pl. Prodanov, A. Drygajlo, "Bayesian networks for error handling through multimodality fusion in spoken dialogues with mobile robots," *Workshop SAPA.04*, Oct. 3, 2004, Jeju, Korea.
[12] Pl. Prodanov, A. Drygajlo, "Bayesian networks for spoken dialogue management in multimodal systems of tour-guide robots," *Proc. of the 8th European Conf. on Speech Communication and Technology*, Eurospeech 2003, Geneva, Switzerland, Sept. 2003, pp. 1057-1060.
[13] St. Russell, P. Norvig, *Artificial Intelligence A Modern Approach*, Prentice Hall, 2003.
[14] M. Turunen, J. Hakulinen, "Agent-based error handling in spoken dialogue systems," In Proc. Eurospeech 2001, pp. 2189-2192.
[15] T. Willeke, C. Kunz, I. Nourbakhsh, "The history of the Mobot museum robot series: An evolutionary study," *FLAIRS* 2001, May, 2001.