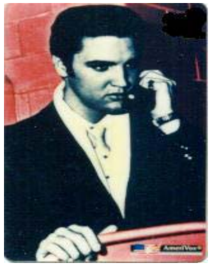
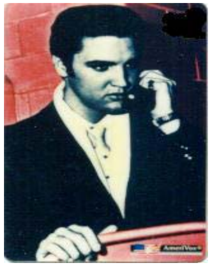
Towards Customizable Individualized Dialogue Systems

Marilyn Walker, S. Whittaker, R. Moore, J. Moore and S. Young
Universities of Sheffield,
Edinburgh and Cambridge

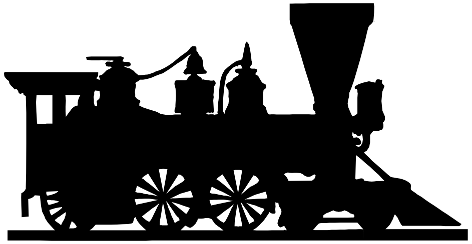


Spoken Dialogue Systems (HCI and Human Modeling)

- An intelligent artifact that can interact with humans to complete certain tasks
- An important experimental vehicle for Cognitive Science
- Cognitive hypotheses about dialogue can be embodied in system and tested
- Hypotheses related to individual differences in interaction



Spoken Dialogue Systems: THE PAST to the PRESENT

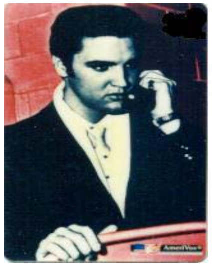


- Travel information systems, e.g. ATIS, SUNDIAL, Communicator
- System initiative limited vocabulary dialogue => mixed initiative large vocabulary ASR
- Commercial systems in many domains (but still limitations)

Key of “Mr. Right” is **instructability** and **feedback**

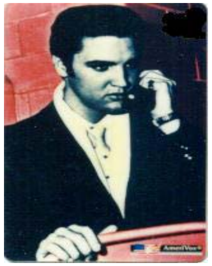
Systems that are **personalized** to respect individual differences

That **learn** over time how to improve their performance

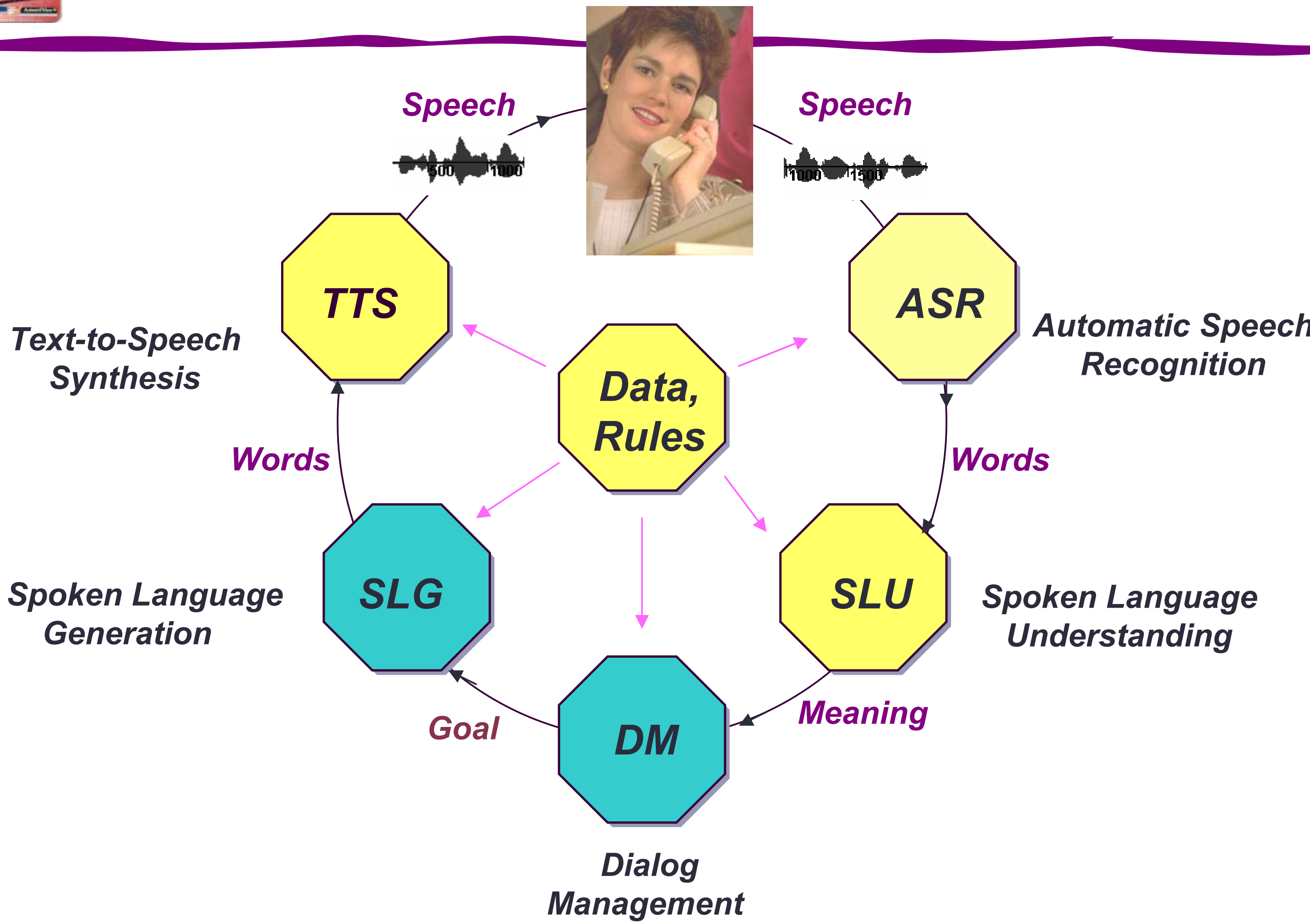


What kind of learning would be important?

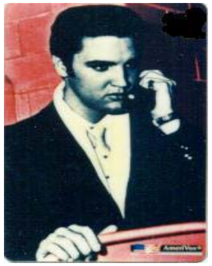
- ***Modeling individual differences***
- One of the most consistent results from cognitive science
  - Cognitive load (young vs. the elderly)
  - Learning differences (visual vs. verbal learners)
  - Interactive style (casual vs. formal, directive vs. nondirective)



Spoken Dialogue Systems

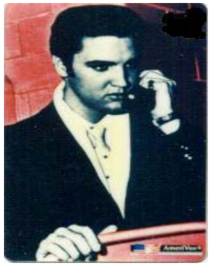


Hypothesis: Individualization and customization depend on methods for training the DM and SLG



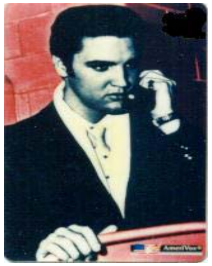
Training the Dialogue Manager and Response Generator

- Dialogue management:
  - Reinforcement learning: Levin etal 97, Walker etal 98, Litman etal, 2000, Scheffler and Young 2002
- Spoken language generation
  - Decision-theoretic user models (Carenini and Moore IJCAI 2001a, b; Walker etal, Cognitive Science 2004)
  - Rankboost (form of boosting) (Rambow etal ACL01, Walker etal NAACL01, Stent etal ACL04)



Reinforcement Learning

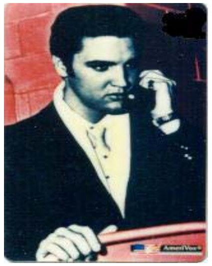
- System characterized in terms of a set of states, and actions that can be taken in each state
- Actions can be something said to the user, a whole subdialogue, or accessing the database
- The rewards received on reaching a state or at end of dialogue are used to learn which actions lead to highest rewards



Reinforcement Learning (cont)

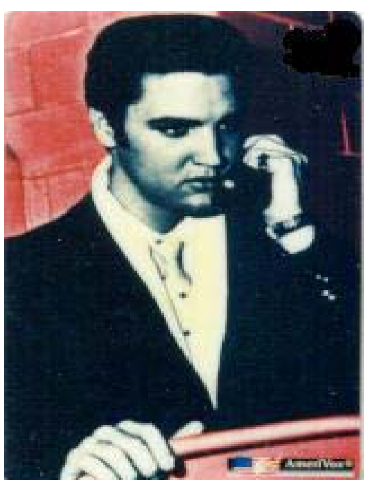
$$U(a, S_i) = R(S_i) + \sum_j M_{ij}^a \max_{a'} U(a', S_j)$$

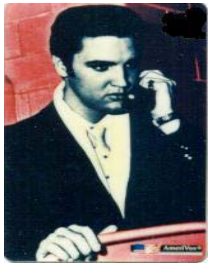
- Actions a, a’
- States Si, Sj
- Utility U
- M^a_{ij} , probability of going from State i to State j on doing action a (estimated from experimental data)
- R(Si) - the immediate reward for getting to State i
- U(final state): the delayed reward for completing the dialogue



Experiments with human users

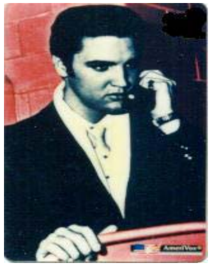
- ELVIS: User Satisfaction increased from 27.5 (training) to 31.7 (test) ($p < .05$)
- NJFun: Task Completion increased from .52 (training) to .64 (test) ($p < .06$)





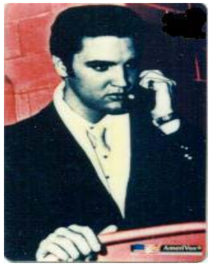
Individualizing the reward function?

- Cognitive load (young vs. the elderly)
- Learning differences (visual vs. verbal learners)
- Interactive style (casual vs. formal, directive vs. nondirective)



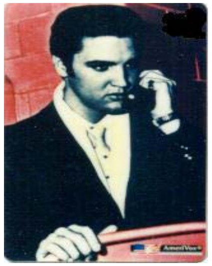
Summary (RL)

- Reinforcement learning allows you to represent any system actions as choices the system is making in a particular dialogue state
- The reward function can be based on any evaluation metric you wish to optimize
- Experiments so far suggest that the method provides measurable and significant system improvements on chosen metric



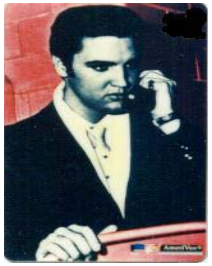
Decision-theoretic models for content selection

- User interacts with system to indicate importance of different domain attributes in decision making
- User-tailored responses select content depending on individual preferences
- Experiments in real-estate, restaurant and travel domains show increased effectiveness in decision making (Carenini and Moore, Walker etal, Stent etal, Moore etal)
- Open questions about degree of conciseness and form of information presentation (cognitive load for processing information)



Boosting to customize form of response for SLG

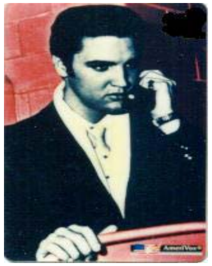
- Example responses represented by a set of features describing any potential aspect of the response
- Each response has an associated ***rating*** derived from ***human feedback*** (e.g. Informational Coherence)
- These ratings induce a partial order over the set of possible responses
- The training method learns how to reproduce this partial order (ranking) of responses.



Rankboost Algorithm (See Schapire 99, Iyer etal 98)

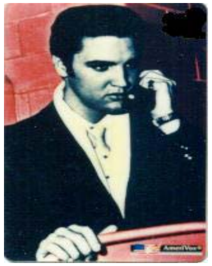
- Each response x represented as sum of *m* indicator functions where each function threshholded on a feature count:

  $h_s(x) = 1$ *if feature-count* > 1, *else* 0
- Each function $h_s(x)$ has single α_s parameter
- Ranking Score: $\mathbf{F(x)} = \Sigma_s \alpha_s h_s(x)$, ranks competing responses
- Training data is a set of pairs (x,y) for each example x rated higher than y
- Training: set the parameters α_s to minimize the loss function $Loss = \Sigma_{(x,y)}\, e^{-(F(x) - F(y))}$
- As *Loss* is minimized, $(F(x) - F(y))$ where x is preferred to y is pushed to positive and ranking errors will tend to be reduced



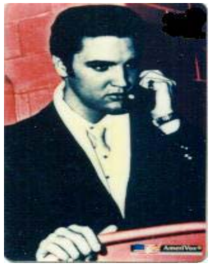
Examples: Learned Rules applied to test fold

Realization	Human	RankBoost
Babbo has the best overall quality among the selected restaurants because it has superb food quality, with excellent service, and it has excellent decor.	1.5	0.45
Babbo has excellent service. It has superb food quality. It has excellent decor. It has the best overall quality among the selected restaurants.	2.0	0.21
Since Babbo has excellent service and superb food quality, with excellent decor, it has the best overall quality among the selected restaurants.	3.5	0.77
Babbo has excellent service and superb food quality, with excellent decor. It has the best overall quality among the selected restaurants	4	0.88
With excellent decor, excellent service and superb food quality, Babbo has the best overall quality among the selected restaurants..	5	0.91



Response Generation Summary

- Can train response generation to rank possible responses
- Example: Ranking based on user feedback on the response's informational coherence
- However, user feedback could be oriented to any evaluation metric associated with the response, e.g. measures collected automatically via neurophysiological probes



Proposal for Individualization

- Reward function for reinforcement learning could be based in individual feedback
- Multi-attribute models individualize content selection but degree of conciseness and form of presentation left uncustomized
- Boosting method would support individualized conciseness and presentation form, given individualized feedback for ranking
- Need research on metrics, probes to collect them, which differences most important, methods for training with smaller amounts of human interactive/feedback data