



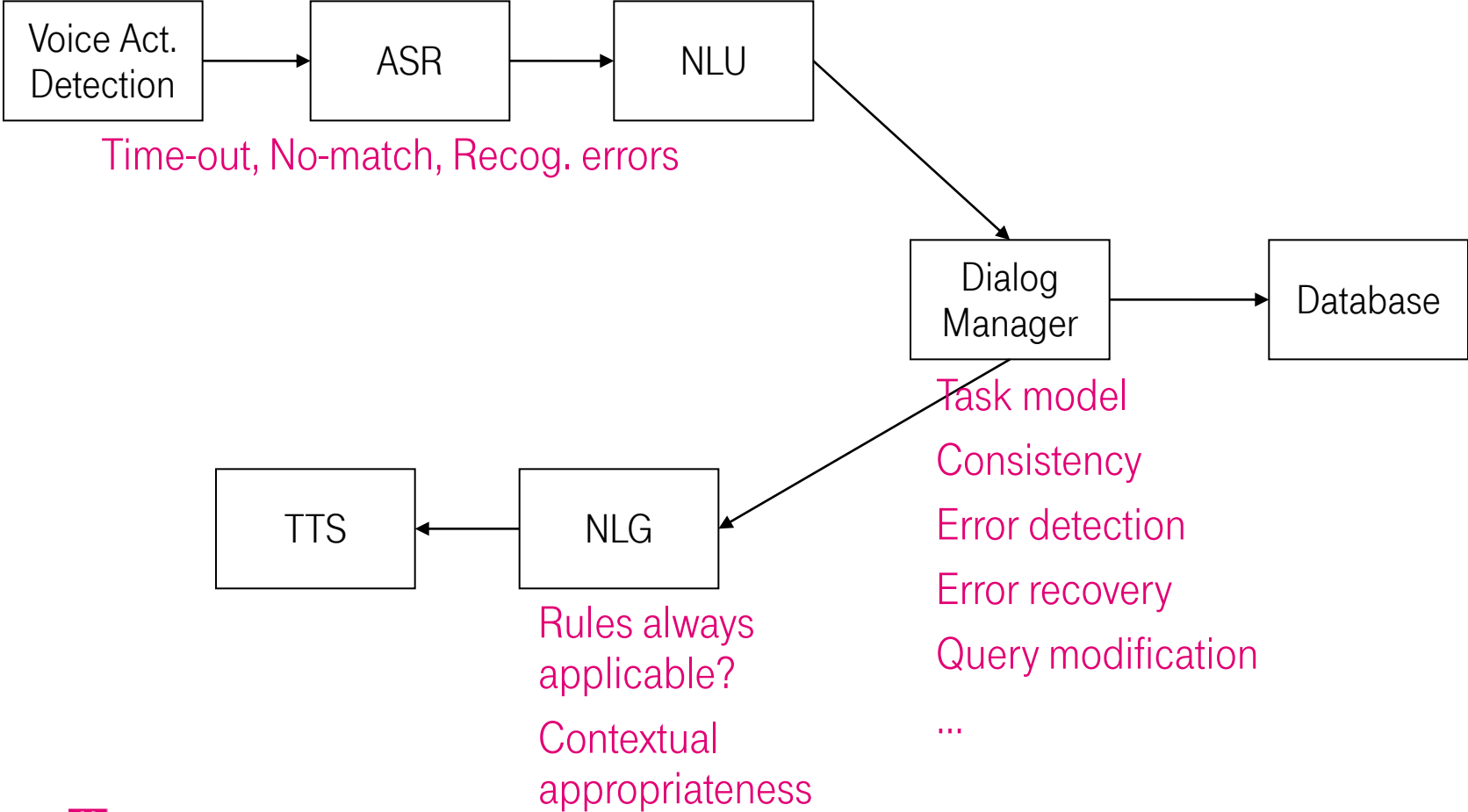
it ▪ Campus ▪ Pioneering ▪ Research ▪ Technologies ▪ Markets ▪ Social Developments ▪ Visions ▪ Future ▪ Design ▪ Seamless Home Environment ▪ Expertise ▪ amless Home Environment ▪ Expertise ▪ Crystallization ▪ Innovation ▪ Development ▪ Processes ▪ BroadWave ▪ Communications ▪ Industrial Partners ▪ Berlin ▪ cations ▪ Industrial Partners ▪ Berlin ▪ Markets ▪ Innovation ▪ T-Com ▪ Universities ▪ Strategy ▪ Market ▪ Trends ▪ Portfolio ▪ Broadband ▪ Virtual City Guide ▪ olio ▪ Broadband ▪ Virtual City Guide ▪ Pervasive Communications ▪ Intuitive Usability ▪ AAA Architecture ▪ Intelligent Access ▪ Inherent Security ▪ Infrastructure cess ▪ Inherent Security ▪ Infrastructure Development ▪ Industry Partners ▪ Information ▪ Scientists ▪ Market Trends ▪ Contextual Information to Go ▪ Entrepreneurs tual Information to Go ▪ Entrepreneurs ▪ Campus ▪ Pioneering ▪ Research ▪ NetShield ▪ Markets ▪ Experts ▪ Visions ▪ Future ▪ Design ▪ Know-how ▪ Expertise ▪ re ▪ Design ▪ Know-how ▪ Expertise ▪ Innovation ▪ Development ▪ Processes ▪ Industry ▪ Communications ▪ Mobile Tracking Device ▪ Berlin ▪ Laboratories ▪ acking Device ▪ Berlin ▪ Laboratories ▪ Research ▪ Innovation ▪ Media Provisioning ▪ Laboratory ▪ Innovation ▪ Customers ▪ Market ▪ Trends ▪ Community-enabling ▪ Market ▪ Trends ▪ Community-enabling Services ▪ Broadband Wireless Access ▪ Pervasive Communications ▪ Intuitive Usability ▪ Integrated Communication ▪ Intelligent ▪ Integrated Communication ▪ Intelligent Access ▪ Personal Intelligent User Interfaces ▪ Network ▪ Information ▪ Gesture-based Real-time Animated Avatars ▪ Market sed Real-time Animated Avatars ▪ Market Trends ▪ New Business ▪ Speech-based Classification ▪ Campus ▪ Pioneering ▪ Research ▪ Technologies ▪ Markets ▪ Affective rch ▪ Technologies ▪ Markets ▪ Affective Interfaces ▪ New Business ▪ Future ▪ Design ▪ Know-how ▪ Expertise ▪ Sensor Nets ▪ Innovation ▪ Processes ▪ Technologies ▪ Innovation ▪ Processes ▪ Technologies ▪ Communications ▪ Berlin ▪ Laboratories ▪ Projects ▪ Innovation ▪ Development ▪ Laboratory ▪ Quality ▪ Strategy ▪ Communications ▪ Berlin ▪ Laboratories ▪ Projects ▪ Innovation ▪ Development ▪ Laboratory ▪ Quality ▪ Strategy ▪ Continuous Sound for Interaction ▪ Trends ▪ Portfolio ▪ Broadband ▪ Creative Potential ▪ Pervasive Communications ▪ Intuitive

Vorhersage von Benutzerurteilen mit Hidden-Markov-Modellen

Klaus-Peter Engelbrecht, Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin



Motivation



Three reasons why we need quality judgments

1) We are interested in the users' **opinion**



Three reasons why we need quality judgments

- 1) We are interested in the users' **opinion**
- 2) They give an idea of the **overall goodness** of the system

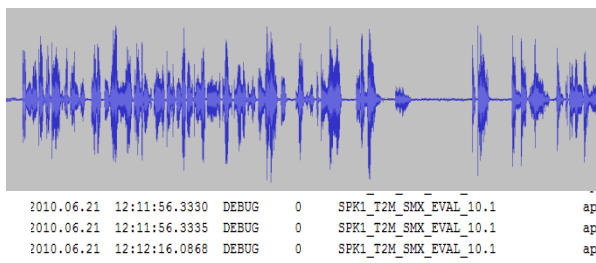


Three reasons why we need quality judgments

- 1) We are interested in the users' **opinion**
- 2) They give an idea of the **overall goodness** of the system
- 3) They help **discover usability problems**



Interaction parameters



```
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::F_GetPromptNumbe
CStateMachine::PerformASR: "cli
.....
```

- Duration: DD, STD, UTD, SRD, URD
- System events: #no-match, #time-out, #barge-in, #help_message)



Transcription of user utterances

- Counts: #Turns, WPST, WPUT, #failed_barge-in
- ASR performance (WER)



Annotation

- Semantic concepts
- Task Success

- Concept level: #AVPs, NLU performance (CER)
- Task Success
- Derivatives: Errors per Turns, #AVPs per Turns



A more fine-grained model of the judgment process

	Error	Open.	...
Welcome. When would you like to eat?	PA:CO	2	...
Today in the evening			
I didn't understand. When do you want to eat?	PA:FA	2	...
Today in the evening.			
At which time today do you want to eat?	PA:PA	1	...
Evening.			
OK, tonight. In which district?	PA:CO	1	...
Wattenscheid.			
OK, Wattenscheid. Which cuisine?	PACO	1	...
Italian.			
I couldn't understand you. Please repeat!	PA:FA	1	...
Italian.			
OK, this is your restarant: ...	PA:CO	n.a.	...



A more fine-grained model of the judgment process

	Error	Open.	...	Judgment U_1
Welcome. When would you like to eat?	PA:CO	2	...	„good“
Today in the evening I didn't understand. When do you want to eat?	PA:FA	2	...	↓ „bad“
Today in the evening. At which time today do you want to eat?	PA:PA	1	...	→ „bad“
Evening. OK, tonight. In which district?	PA:CO	1	...	→ „bad“
Wattenscheid. OK, Wattenscheid. Which cuisine?	PACO	1	...	↑ „fair“
Italian. I couldn't understand you. Please repeat!	PA:FA	1	...	↓ „bad“
Italian. OK, this is your restarant: ...	PA:CO	n.a.	...	↓ „very bad“



A more fine-grained model of the judgment process

	Error	Open.	...	Judgment U_1	Judgment U_2
Welcome. When would you like to eat?	PA:CO	2	...	„good“	„good“
Today in the evening I didn't understand. When do you want to eat?	PA:FA	2	...	↓ „bad“	↓ „fair“
Today in the evening. At which time today do you want to eat?	PA:PA	1	...	→ „bad“	↓ „bad“
Evening. OK, tonight. In which district?	PA:CO	1	...	→ „bad“	→ „bad“
Wattenscheid. OK, Wattenscheid. Which cuisine?	PACO	1	...	↑ „fair“	→ „bad“
Italian. I couldn't understand you. Please repeat!	PA:FA	1	...	↓ „bad“	↓ „very bad“
Italian. OK, this is your restarant: ...	PA:CO	n.a.	...	↓ „very bad“	↑ „fair“



A more fine-grained model of the judgment process

Welcome. When would you like to eat?

Today in the evening

I didn't understand. When do you want to eat?

Today in the evening.

At which time today do you want to eat?

Evening.

OK, tonight. In which district?

Wattenscheid.

OK, Wattenscheid. Which cuisine?

Italian.

I couldn't understand you. Please repeat!

Italian.

OK, this is your restarant: ...

Error Open. ...

PA:CO 2 ...

PA:FA 2 ...

PA:PA 1 ...

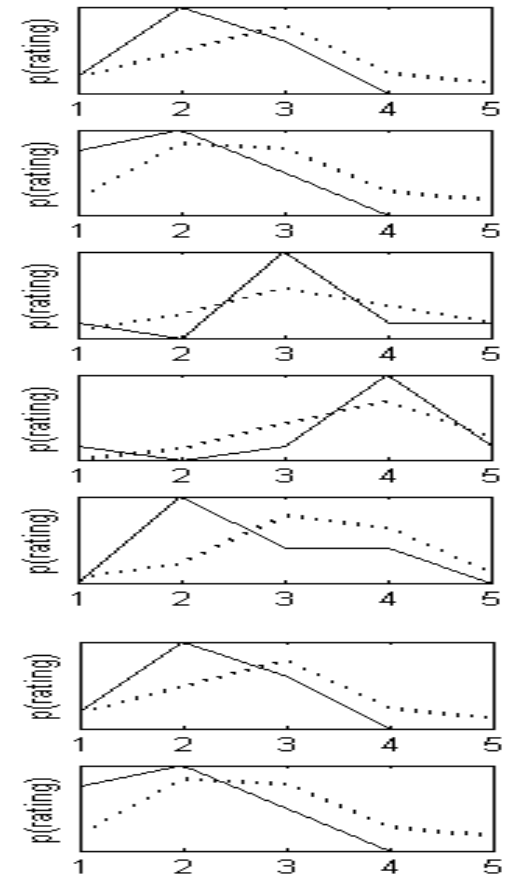
PA:CO 1 ...

PACO 1 ...

PA:FA 1 ...

PA:CO n.a. ...

Judgments $U_1 \dots U_N$



A more fine-grained model of the judgment process

Welcome. When would you like to eat?

Today in the evening

I didn't understand. When do you want to eat?

Today in the evening.

At which time today do you want to eat?

Evening.

OK, tonight. In which district?

Wattenscheid.

OK, Wattenscheid. Which cuisine?

Italian.

I couldn't understand you. Please repeat!

Italian.

OK, this is your restarant: ...

Error Open. ...

PA:CO 2 ...

PA:FA 2 ...

PA:PA 1 ...

PA:CO 1 ...

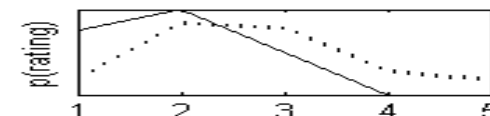
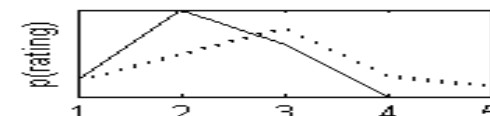
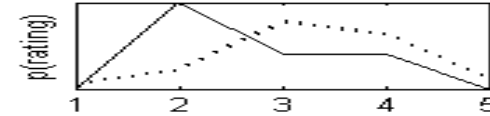
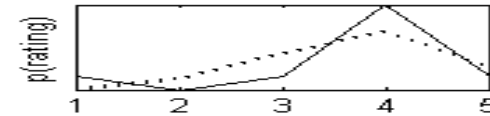
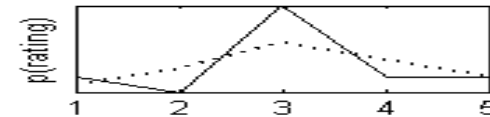
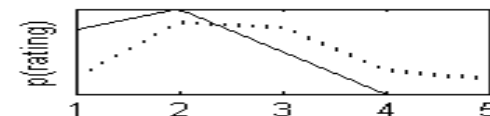
PACO 1 ...

PA:FA 1 ...

PA:CO n.a. ...

Judgments $U_1 \dots U_N$

$$P(J_1 | \text{Error}_1, \dots) * P(J_1)$$



A more fine-grained model of the judgment process

Welcome. When would you like to eat?

Today in the evening

I didn't understand. When do you want to eat?

Today in the evening.

At which time today do you want to eat?

Evening.

OK, tonight. In which district?

Wattenscheid.

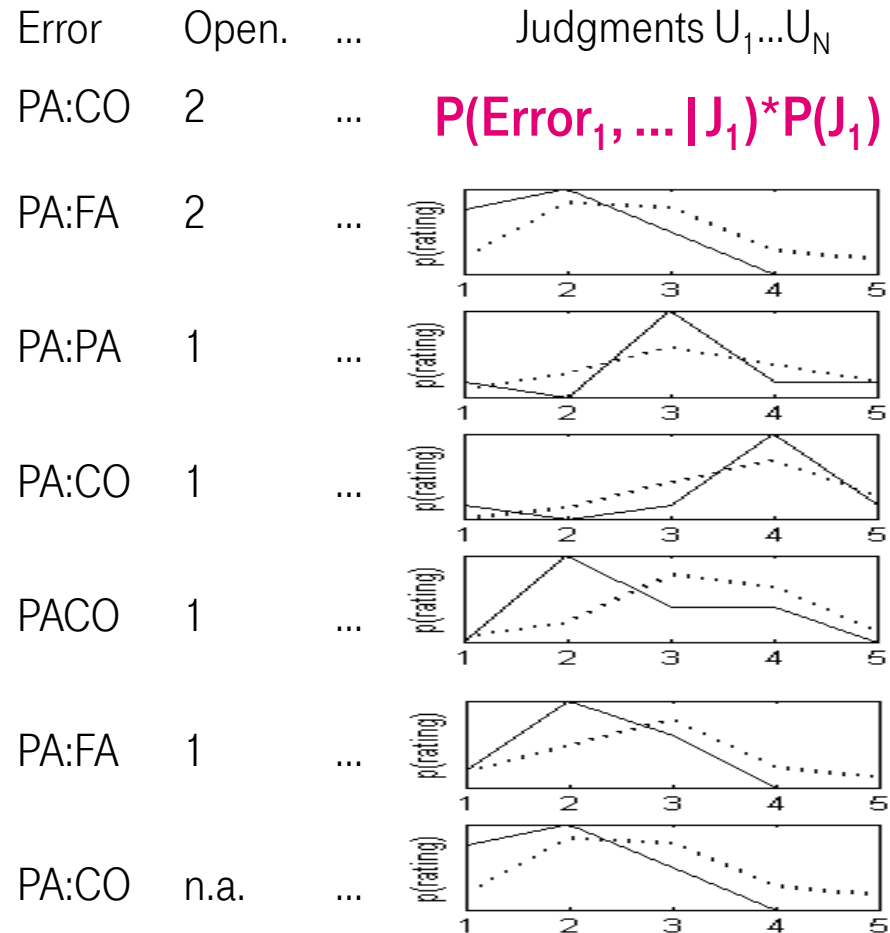
OK, Wattenscheid. Which cuisine?

Italian.

I couldn't understand you. Please repeat!

Italian.

OK, this is your restarant: ...



A more fine-grained model of the judgment process

Welcome. When would you like to eat?

Today in the evening

I didn't understand. When do you want to eat?

Today in the evening.

At which time today do you want to eat?

Evening.

OK, tonight. In which district?

Wattenscheid.

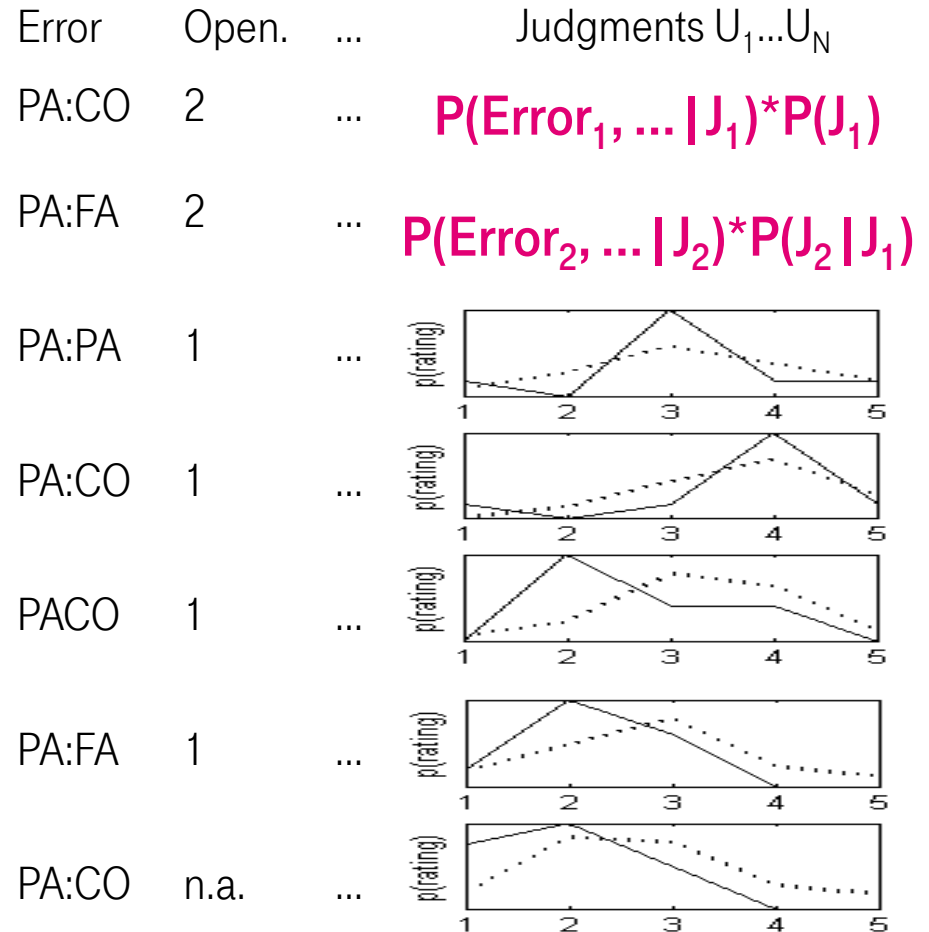
OK, Wattenscheid. Which cuisine?

Italian.

I couldn't understand you. Please repeat!

Italian.

OK, this is your restaurant: ...

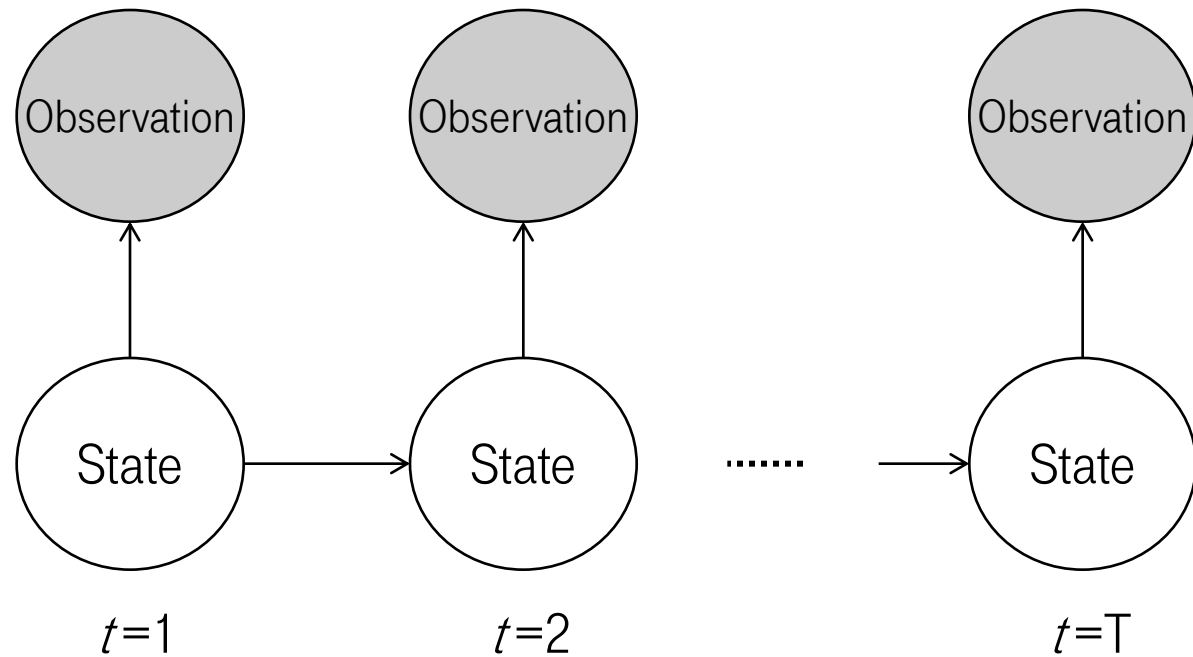


A more fine-grained model of the judgment process

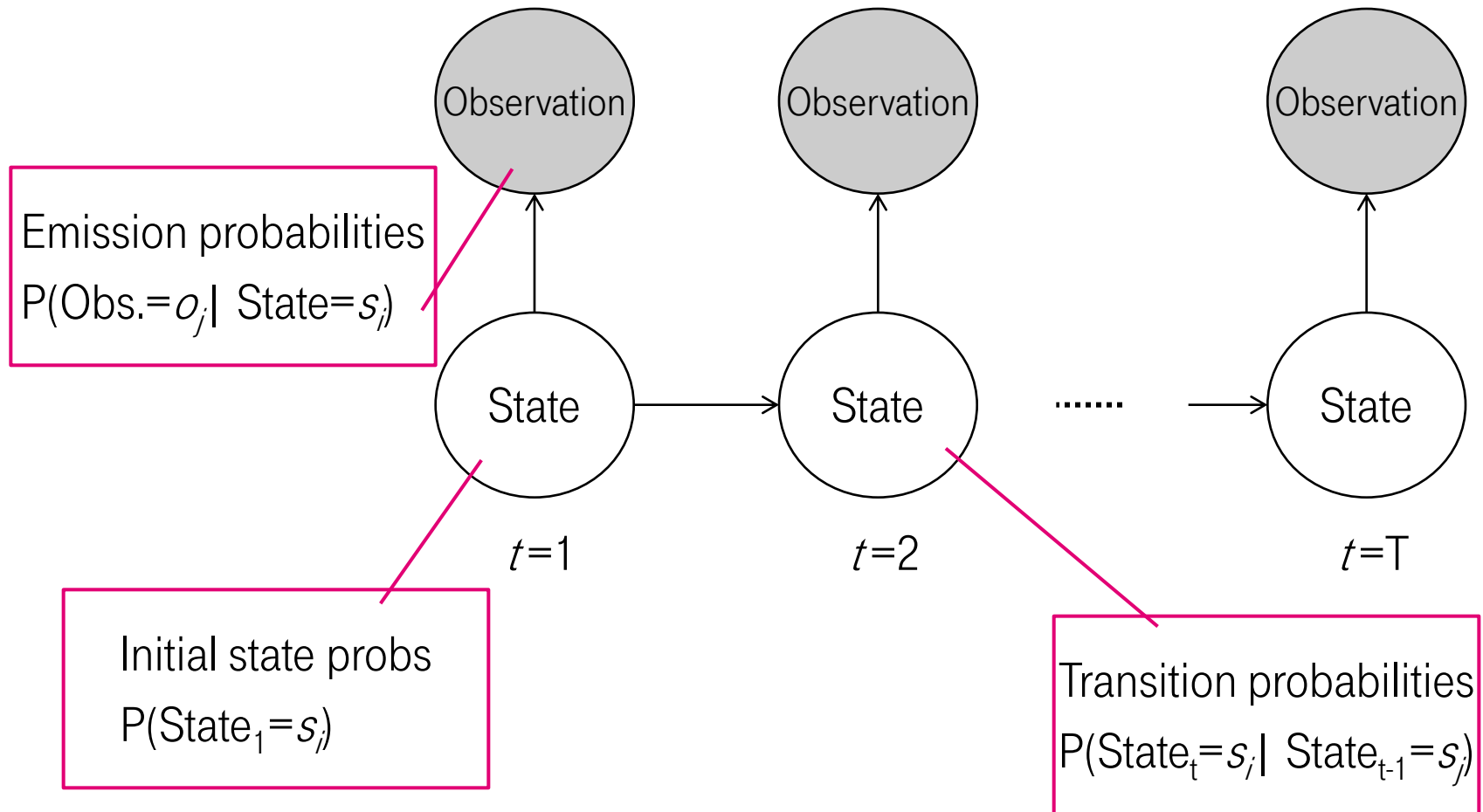
	Error	Open.	...	Judgments $U_1 \dots U_N$
Welcome. When would you like to eat?	PA:CO	2	...	$P(\text{Error}_1, \dots J_1) * P(J_1)$
Today in the evening				
I didn't understand. When do you want to eat?	PA:FA	2	...	$P(\text{Error}_2, \dots J_2) * P(J_2 J_1)$
Today in the evening.				
At which time today do you want to eat?	PA:PA	1	...	$P(\text{Error}_3, \dots J_3) * P(J_3 J_2)$
Evening.				
OK, tonight. In which district?	PA:CO	1
Wattenscheid.				
OK, Wattenscheid. Which cuisine?	PACO	1	...	
Italian.				
I couldn't understand you. Please repeat!	PA:FA	1	...	
Italian.				
OK, this is your restarant: ...	PA:CO	n.a.	...	



Hidden-Markov-Model



Hidden-Markov-Model

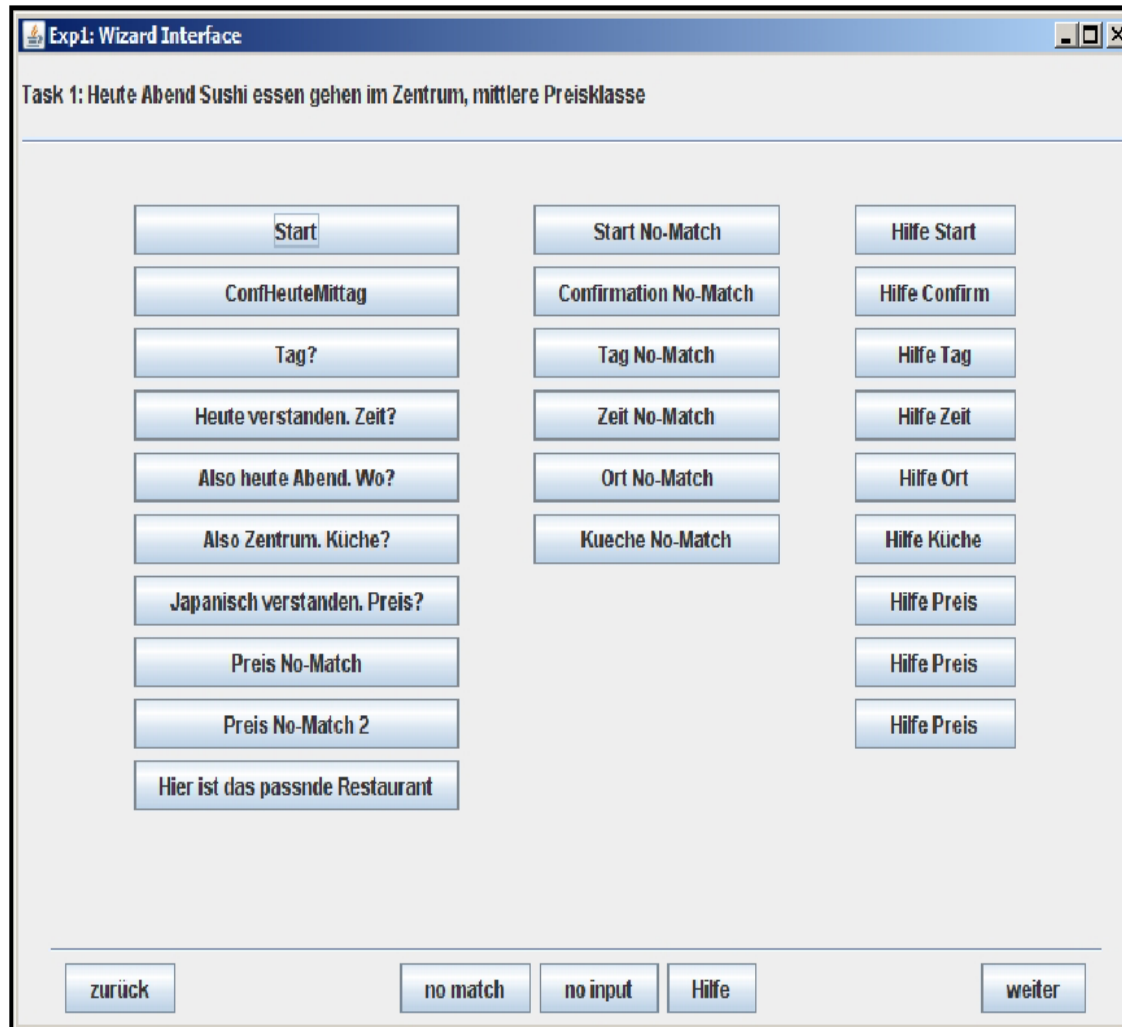


A more fine-grained model of the judgment process

	Error	Open.	...	Judgment U ₁	Judgment U ₂
Welcome. When would you like to eat?	PA:CO	2	...	„good“	„good“
Today in the evening I didn't understand. When do you want to eat?	PA:FA	2	...	↓ „bad“	↓ „fair“
Today in the evening. At which time today do you want to eat?	PA:PA	1	...	→ „bad“	↓ „bad“
Evening. OK, tonight. In which district?	PA:CO	1	...	→ „bad“	→ „bad“
Wattenscheid. OK, Wattenscheid. Which cuisine?	PACO	1	...	↑ „fair“	→ „bad“
Italian. I couldn't understand you. Please repeat!	PA:FA	1	...	↓ „bad“	↓ „very bad“
Italian. OK, this is your restarant: ...	PA:CO	n.a.	...	↓ „very bad“	↑ „fair“



Wizard-of-Oz Interface



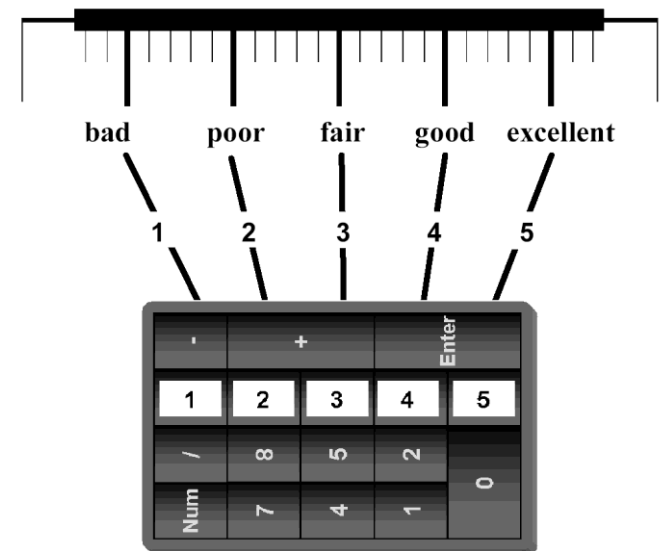
The experiment from the user's point of view

Aufgabe 1

Stellen Sie sich vor, sie sind **heute Abend** als Tourist in Bochum, um sich ein Musical anzusehen. Bevor die Aufführung beginnt, möchten Sie noch **Sushi** essen gehen. Sie rufen bei einem automatischen Auskunftssystem an, um ein Sushi-Restaurant im **Zentrum** zu suchen. Da die Karten teuer waren, möchten Sie ein Restaurant in der mittleren Preisklasse.

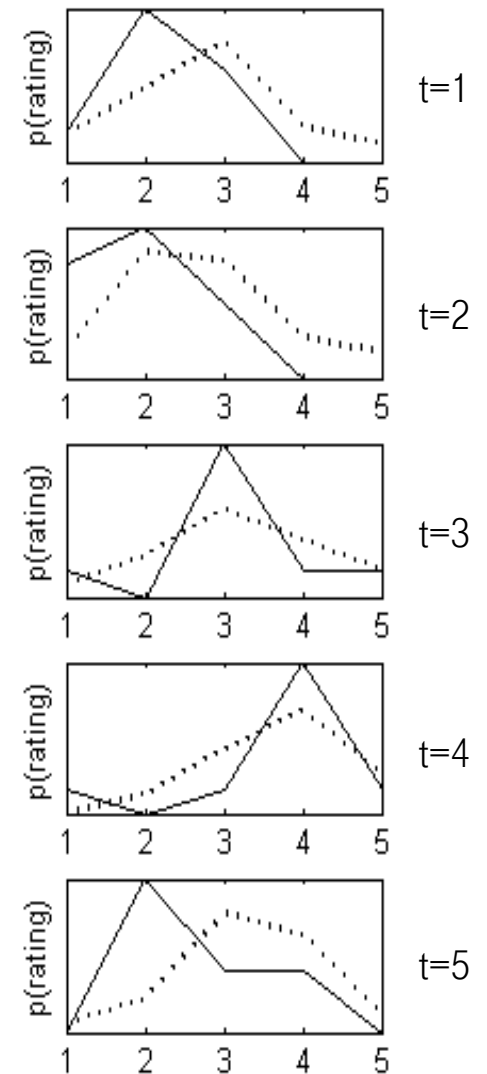
- Heute Abend
- Zentrum
- Sushi
- Mittlere Preisklasse

**Please rate the quality of the dialog
up to the current moment!**



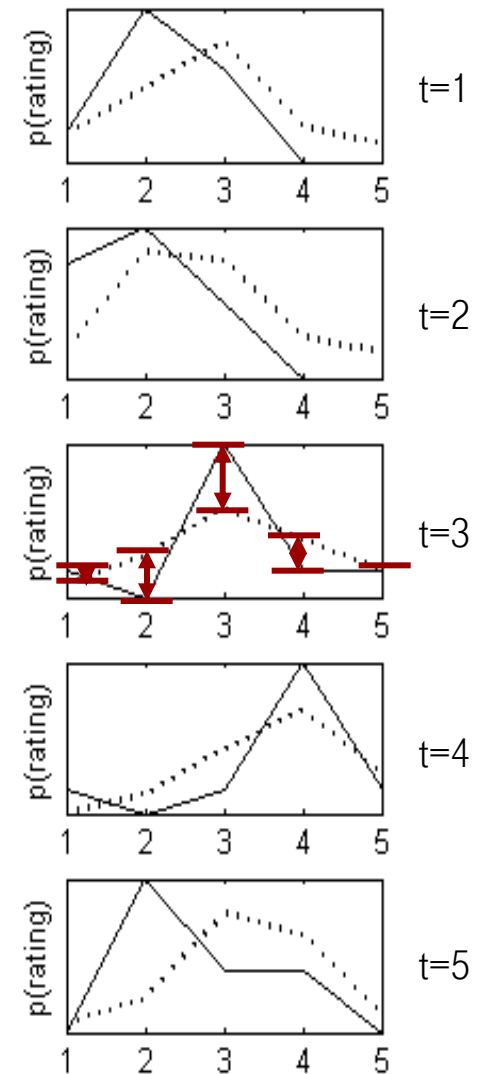
Model performance

- MSE_{dist} : differences in probabilities for judgments 1...5 averaged over all judgments in all turns



Model performance

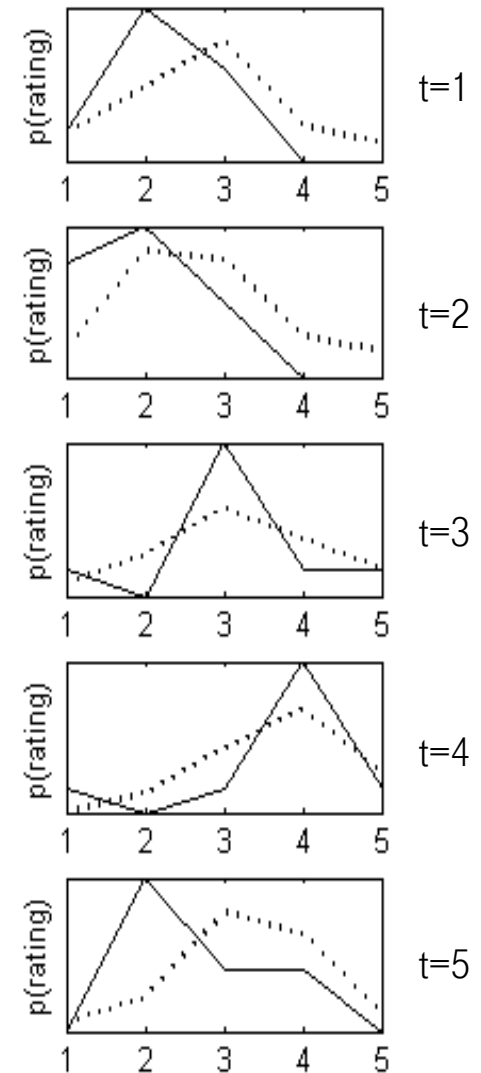
- MSE_{dist} : differences in probabilities for judgments 1...5 averaged over all judgments in all turns



Model performance

- MSE_{dist} : differences in probabilities for judgments 1...5 averaged over all judgments in all turns

		Lin. Reg.	HMM
Last turn	Train	0.035	0.054
	L1O	0.069	0.078
	Test	0.225	0.055
All turns	Train	0.059	0.035
	L1O	0.081	0.052
	Test	0.130	0.044



Summary

- **Overall quality** of the system from the **users' perspective**



Summary

- **Overall quality** of the system from the **users' perspective**
- Judging a dialog system as **continuous process**



Summary

- **Overall quality** of the system from the **users' perspective**
- Judging a dialog system as **continuous process**
- Can be modeled with **Hidden-Markov-Models**

