



Adaptation of an automotive dialogue system to users' expertise and evaluation of the system

Dragan Milchevski

Seminar: Personalization of User Interface

Supervisor: Michael Feld

University of Saarland

Google Glass Project



Motivation



[1]



[2]

Motivation



Computer, show me
the shortest route
to Saarbrücken

[3]



[4]

[5]



History

- First in-car spoken dialogue systems (SDSs) in late 90s
- Ways of adaptations:
 - The task demand of the user attention
 - Influence of age on user behavior
 - User preferences
 - Clarification of dialogues

User experience



[6]

Inexperienced

Experienced



[7]

Adapting techniques

- Adapting the dialogue strategy
- Adapting functional range
- Adapting the content of the system prompts
 - Lack of research on this issue
 - Traffic must be at the forefront of drivers' awareness
 - Speech dialogues in automotive are very restricted



Outline

- Motivation
- Classification of the users
- Adaptation of user prompts
- Evaluation
- Related work
- Conclusion

Classification of the users



[6]

- Accomplish tasks quickly –
- Little or no guidance –
- Short or no confirmation prompts –

Expert

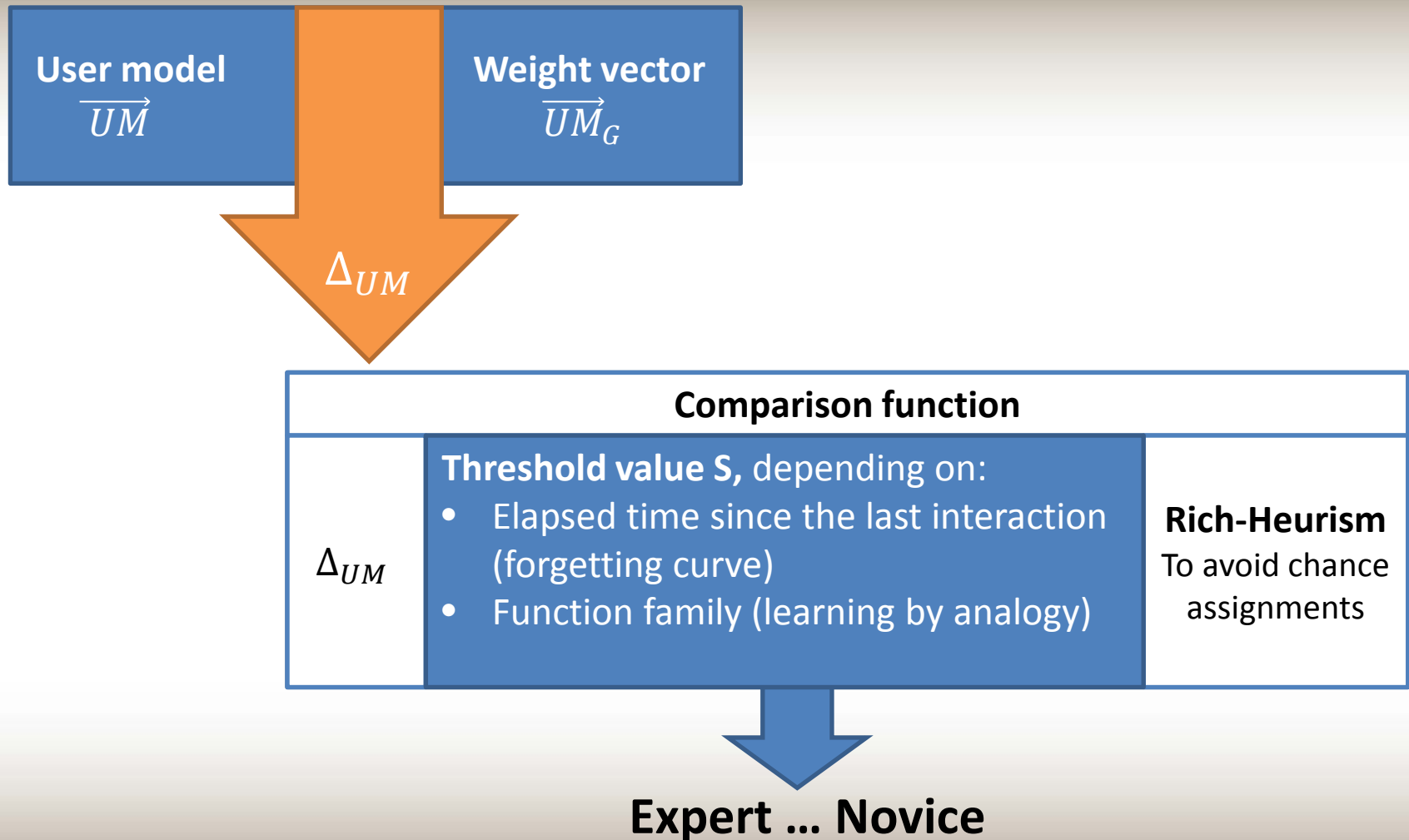
Novice

- Tutorial
- Detail confirmation of the actions



[7]

Calculating the user category



User Model

- Vector \overrightarrow{UM}
- Parameters (h, o, t, e, T_m)
- Calculated for each task separately

Parameter	Meaning
# Help request h	Users asked for general information about the system
# Option requests, o	Users asked for the currently available voice commands
# Timeouts, t	The ASR did not get any acoustic signal (>5 s pause)
# ASR-failure, e	The system could not understand the users' input, e.g. OOV words or unintelligible speech
Onset time, T_m	Users need more than 3 s to start answering

Table 1 Parameters for calculation of \overrightarrow{UM}

Calculation of User Model - example



〈PTT Button〉

$$h = 0$$

$$o = 0$$

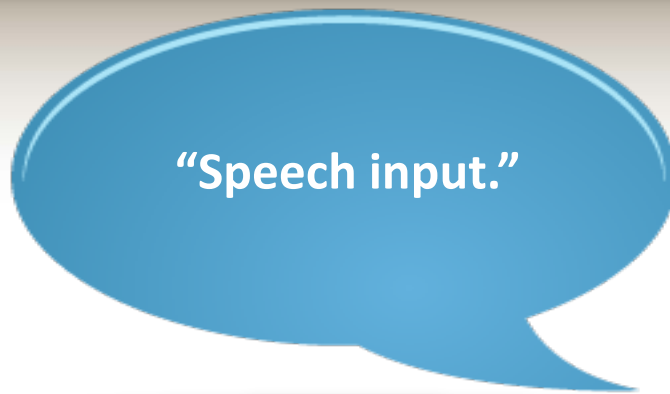
$$t = 0$$

$$e = 0$$

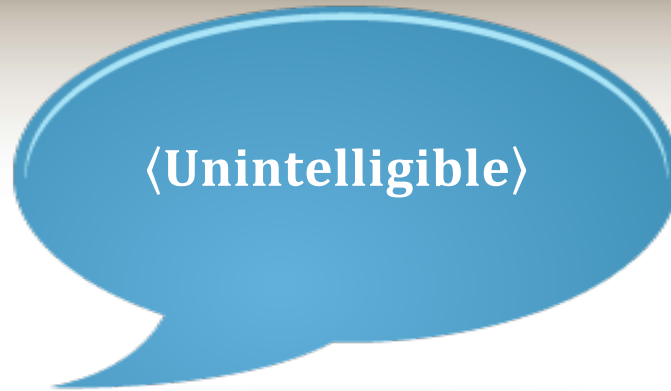
$$T_m = 0$$



Calculation of User Model - example



Calculation of User Model - example



(1) $e = e + 1$



Calculation of User Model - example



"Pardon me?"

(1) $e = e + 1$



Calculation of User Model - example



⟨Unintelligible⟩

$$(1) e = e + 1$$

$$(2) e = e + 1, T_m = T_m + 1$$



Calculation of User Model - example



"I couldn't understand you. Choose a Radio Frequency, Station or.."

$$(1) e = e + 1$$

$$(2) e = e + 1, T_m = T_m + 1$$



Calculation of User Model - example



⟨Unintelligible⟩

- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$



Calculation of User Model - example



"I couldn't understand you. Speech input turned off."

- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$



Calculation of User Model - example

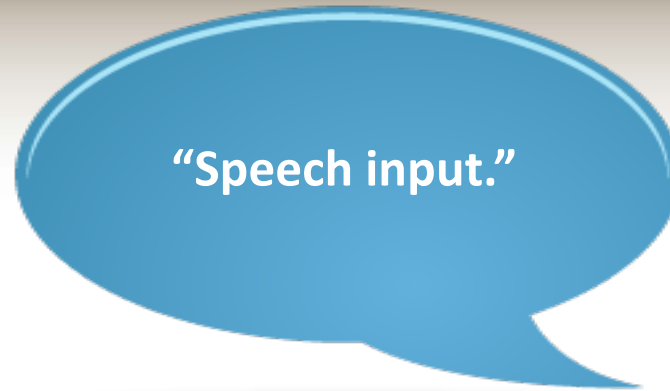


〈PTT Button〉

- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$



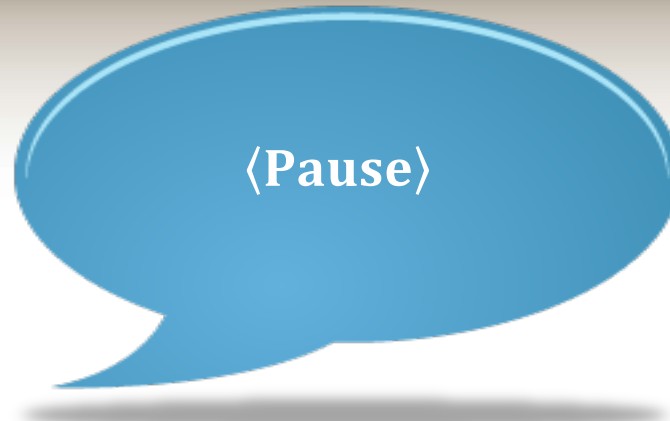
Calculation of User Model - example



- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$



Calculation of User Model - example



- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$



Calculation of User Model - example

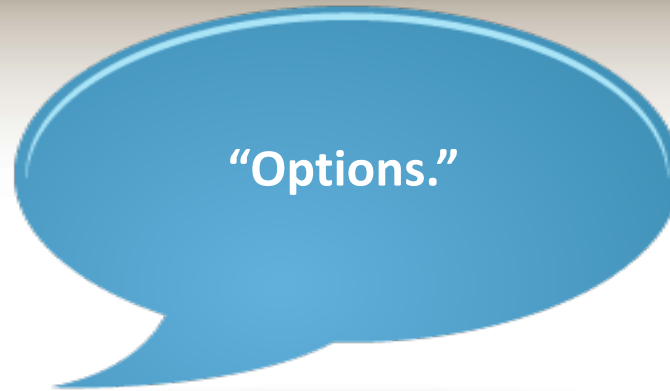


"I couldn't hear you.
Please repeat."

- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$



Calculation of User Model - example



- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$
- (5) $o = o + 1$

Calculation of User Model - example



“You can say: Choose
Frequency, choose
Station...”

- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$
- (5) $o = o + 1$



Calculation of User Model - example



“Choose
Frequency.”



- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$
- (5) $o = o + 1$
- (6) $T_m = T_m + 1$

Calculation of User Model - example



“Which frequency do you want?”



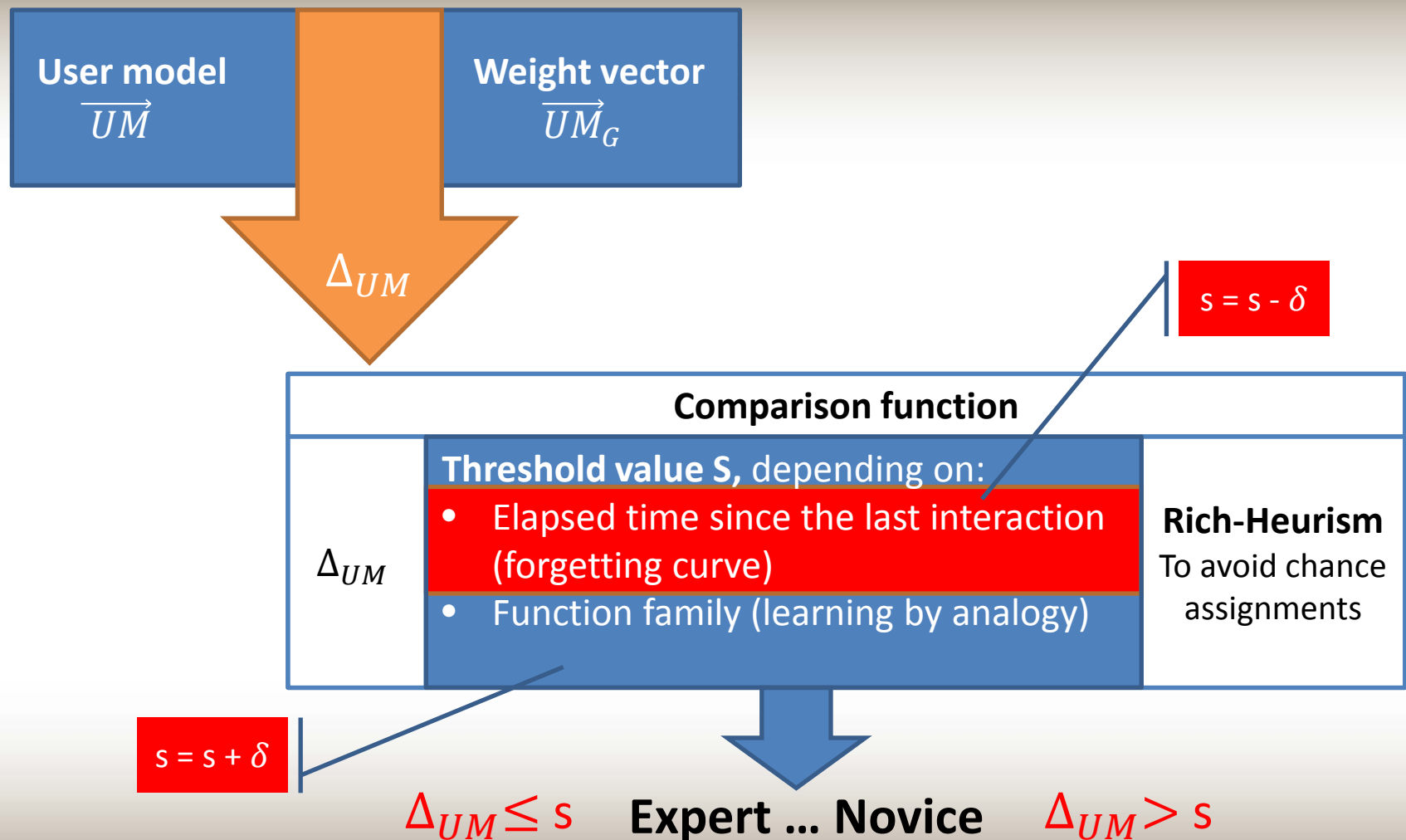
- (1) $e = e + 1$
- (2) $e = e + 1, T_m = T_m + 1$
- (3) $e = e + 1$
- (4) $t = t + 1$
- (5) $o = o + 1$
- (6) $T_m = T_m + 1$

$$\overrightarrow{UM}_{\text{Choose Frequency}} = [h, o, t, e, T_m] = [0, 1, 1, 3, 2]$$

Weight vector

- \overrightarrow{UM}_G represents \overrightarrow{UM} by weights to tell the user experience
- Overall score (User status) is the sum of weighted values
- $\sum_{i=1}^n v_{G_i} = 1$ for each v_{G_i} in \overrightarrow{UM}_G
 - initially $v_{G_i} = \frac{1}{n}$
- Multi linear regression of:
 - The user status as independent variable
 - Recorded components of \overrightarrow{UM} as independent variable^[16] (Hassel 2006)
- Example:
 - $\overrightarrow{UM}_G = \langle v_{G_h} = 0.17 \ v_{G_o} = 0.22 \ v_{G_t} = 0.16 \ v_{G_e} = 0.45 \rangle$ – reference system
 - $\overrightarrow{UM}_G = \langle v_{G_h} = 0.12 \ v_{G_o} = 0.25 \ v_{G_t} = 0.13 \ v_{G_e} = 0.50 \rangle$ - prototype

Calculating the user category



Outline

- Motivation
- Classification of the users
- Adaptation of user prompts
- Evaluation
- Related work
- Conclusion

Adaptation of the users prompts

- Explicitness

Utterance type	Novice	Expert
Opening/closing	U: <PTT> (<i>Action-directive</i>) S: Speech input <Tone A>/Speech input terminated <Tone B>	U: <PTT> S: <Tone A>/<Tone B>
Signal understanding	U: Play CD. (<i>Action-directive</i>) S: CD is being played.	U: Play CD. S: <Music is heard>
Signal understanding (+ Open-option)	U: Entertainment. (<i>Action-directive</i>) S: Entertainment. Say AM, FM, CD or DVD	U: Entertainment. S: Entertainment.
Assert	U: Destination input. (<i>Action-directive</i>) S: <i>This task is currently not available</i>	U: Destination input S: Currently not available.
(Assert +) Action-directive	U: Select CD. (<i>Action-directive</i>) S: CD slot is empty. Insert a CD.	U: Select CD. S: Insert CD.
(Signal non-understanding +) Action-directive	U: <Not recognized> S: I could not understand you, repeat.	U: <Not recognized> S: Pardon me?

Outline

- Motivation
- Classification of the users
- Adaptation of user prompts
- Evaluation
- Related work
- Conclusion

Evaluation – test design



[8]

BMW 7 Series



[9]

BMW 5 Series

Prototype

Reference system



Evaluation – test design

- BMW iDrive
- PARADISE evaluation framework
- Real driving conditions
- 30 males and 14 females
- 22 subjects for each system
- Driving part (30-40 min) and questionnaire
- No one has used the systems before

BMW iDrive

- Keep track of all important things
- Intuitive way of control
- Since 2001
- Configured for 3000 words and phrases
- Available in several languages



PARADISE evaluation framework

- Maximizing the tasks performance
 - Maximize the task success (Kappa coefficient - k)
 - Minimizing the cost (c)
- User satisfaction
 - $US = (\alpha \times N(k)) - \sum_{i=1}^n w_i \times N(c_i)$
 - w_i and α are model parameters
 - $N()$ normalizing function



Evaluation – test tasks

- Task 1: Choose frequency 93.3
 - Task 2: Choose station Bayern 5
 - Task 3: Play title number 4 of the current cd
 - Task 4: Activate traffic program
 - Task 5: Dial a phone number
 - Task 6: Dial a name from the address book
 - Task 7: Display the navigation map
 - Task 8: Change the map scale to 100m
 - Task 9: Change the map style (north, driving, arrows)
 - Task 10: Choose an arbitrary frequency
 - Task 11: Choose an arbitrary station
-

Evaluation - prototype vs. reference

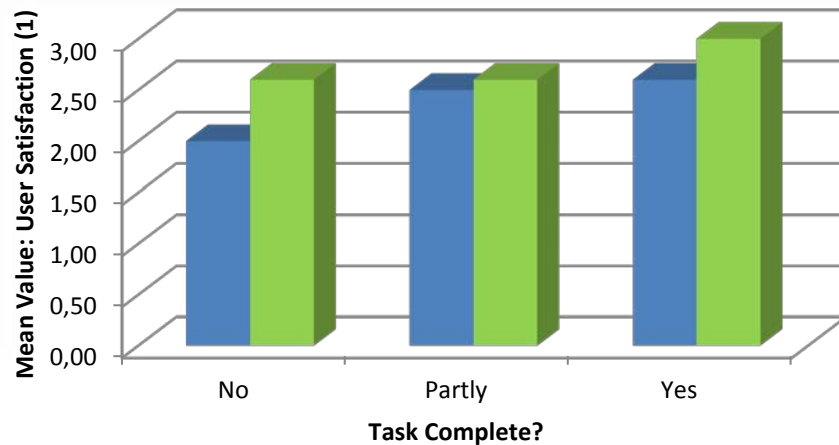
Measure	Prototype	Reference System
ASR-failures per task	1.2 times	1.6 times
Timeouts	2 times more	/
Requested options and help	/	2.5 times more
Commands uttered too soon	12.5 times more	/
Turns to complete task	6.8 turns	8.4 turns
Average time to complete task	40.7 s	60.4 s



Evaluation – user satisfaction

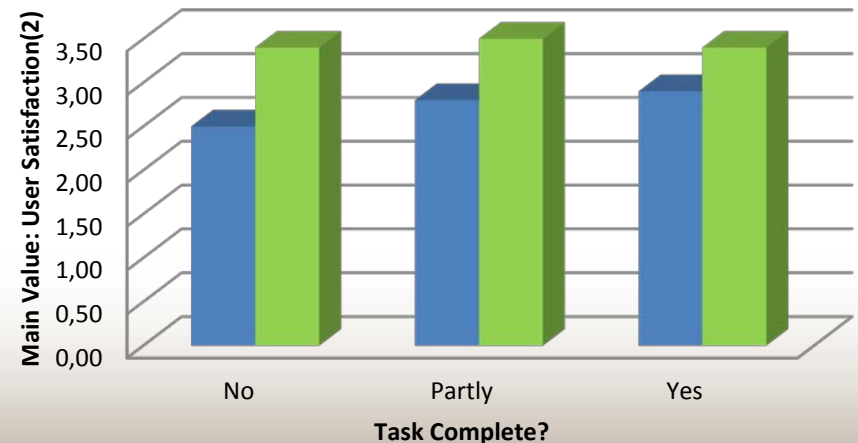
- 6 questions with grades (0 lowest, 4 highest)

■ Reference System (A) ■ Prototype (B)



- I could complete all tasks without problems?
- I find the system easy to use?
- I got frequently upset during the test?

■ Reference System (A) ■ Prototype(B)



- I would recommend the system”
- I really want to have such a system in my car?
- I find the system very useful?

Outline

- Motivation
- Classification of the users
- Adaptation of user prompts
- Evaluation
- Related work
- Conclusion



Virtual Intellectual Co-Driver - VICO

- Robust speech technology
- User- and situation-adaptive intelligent dialogue strategies
- User-friendly, comfortable, efficient and safe-to-use vocal interfaces
- User acceptance and satisfactions

VICO – dialogue example

User: “VICO!”

System: “Yes Please”

User: “I’d like to see the movie ‘What woman want’ tonight.”

System: “In which city would you like to see the movie?”

User: “In Stuttgart”

System: “In Stuttgart ‘What woman want’ is shown at the movie theaters Europa and Gloria. Which one would you prefer?”

User: “That depends on the starting times. I’d like to go at around 8 p.m.”

System: “The Gloria starts at 7.45 p.m., the Europa at 8.15 p.m.”

User: “Then I’d rather go to the Europa theatre.”

System: “Do you want me to make a reservation for you?”

User: “Yes if its possible. I need two tickets.”

System: “One moment please... Two tickets will be reserved for you. They have to be picked up until 8 p.m.”

Outline

- Motivation
- Classification of the users
- Adaptation of user prompts
- Evaluation
- Related work
- Conclusion



Conclusion

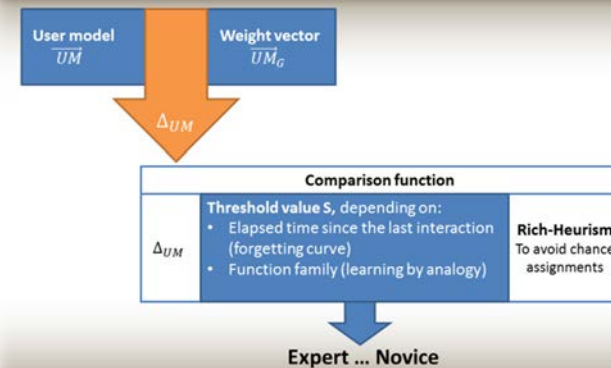
- Adaptation contribute to improve usability
 - Good for learning the system
 - Tedious for long run
- Extended prompts could lead to less straightforward manner of operation
- Better user satisfaction and cost factors

Summary

Classification of the Users



Calculating the User Category

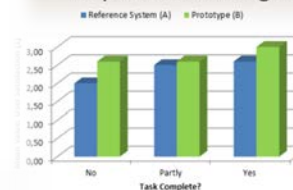


Evaluation – Test Design

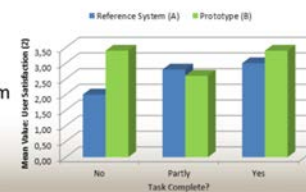


Evaluation – User Satisfaction

- 6 questions with grades (0 lowest, 4 highest)



- I would recommend the system"
- I really want to have such a system in my car?
- I find the system very useful?



Questions



Reference

- [1] http://en.memory-alpha.org/wiki/File:Neural_interface.jpg
- [2] <http://plasma.dk/wordpress/>
- [3] <http://madamenoire.com/wp-content/uploads/2012/05/Woman-driving.jpg>
- [4] <http://images.thetruthaboutcars.com/2012/03/Sanyo-new-car-navigation-system.jpg>
- [5] <http://www.deshow.net/d/file/car/2008-10/bmw-car-z4-2.jpg>
- [6] <http://www.bbpremierins.com/student%20driver%20-%20bb.jpg>
- [7] <http://www.cbc.ca/gfx/images/news/photos/2009/06/19/formula-one-cp-6901673.jpg>
- [8] http://www.buyacar.co.uk/nvdimg/lrg_bmw_7_series_saloon_48161.jpeg
- [9] <http://images.cardekho.com/images/car-images/large/BMW/BMW520d/BMW-5-Series-sedan5.jpg>
- [10] http://www.motorbeam.com/wp-content/uploads/BMW_X1_idrive.jpg
- [11] <http://www.xbimmers.com/forums/attachment.php?attachmentid=199657&stc=1&d=1225491202>
- [12] <http://www.audi-sport.net/vb/a3-s3-sportback-8p-chassis/81540-mobile-telephone-preparation-%96-bluetooth-interface.html>
- [15] http://images.drive.com.au/drive_images/Editorial/2008/10/16/16idrive_m_m.jpg
- [16] Hassel 2006

Backup slides

- Some backup slides follow:

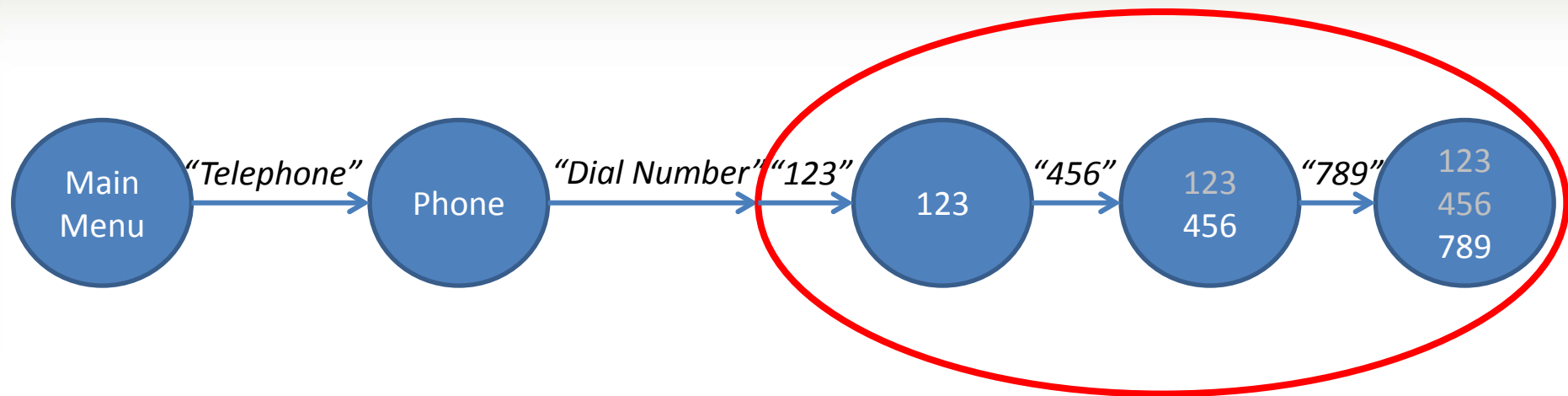
Why did it take the user in the analyzed dialogue about 65 seconds to complete the task?

- Main reason: user is **novice**

User:	< presses PTT >	(1)
System:	"Speech input."	
User:	< Pause >	(2)
System:	"I could not understand you. Please repeat."	
User:	< Speaks too soon> "Change:" < pause > "Oh! Ehh options"	(3)
System:	"Board information". < End of speech input >	
User:	< presses PTT >	(4)
System:	"Speech input"	
User:	"Options"	(5)
System:	"Options. Say board computer, travel board computer ..."	
User:	"Eeeehhh? Map style"	(6)
System:	"Map style. Please choose a style."	
User:	"Options"	(7)
System:	"Options. Choose a map scale, for example: 'scale 5 km', or say: map style north, map style driving or map style arrows"	
User:	"Map style driving"	(8)

Evaluation – Test Task Example

- Dialing a telephone number



[12]