

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Usability Evaluation of Interactive Systems	2
1.2	Automatic Usability Evaluation	4
1.3	Simulation Based Usability Evaluation	5
1.4	Thesis Structure	6
<b>2</b>	<b>State of the Art</b>	9
2.1	Usability Evaluation	10
2.1.1	Methods Supporting Expert Analysis	12
2.1.2	Methods with User Participation	14
2.1.3	Measures in Case of User Participation	15
2.1.4	Model-Based Usability Evaluation	16
2.2	Tools for Automatic Usability Evaluation	22
2.2.1	CogTool	22
2.2.2	CogTool Explorer	22
2.2.3	BisWas	23
2.2.4	SpeechEval	24
2.2.5	MeMo Workbench	24
2.2.6	Selection of the Used Research System	25
2.3	Describing Interactions	26
2.3.1	Methods for the Reconstruction of Practices	26
2.3.2	Classification of User Behaviour	27
2.4	Description and Logging of Multimodal Interactions	29
2.4.1	Design and Development of Multimodal Systems	29
2.4.2	Evaluation of Multimodal Interaction	32
2.4.3	Classification of Description Languages and Modelling Approaches	35

2.5	Measures for the Evaluation of User Behaviour Simulations . . . . .	35
2.5.1	Turn-Level Metrics . . . . .	37
2.5.2	Dialogue-Level Metrics . . . . .	37
2.5.3	Dialogue Act $n$ -grams and Kullback-Leibler Divergence . . .	38
2.6	Research Questions . . . . .	39
<b>3</b>	<b>MeMo Platform for Automatic Usability Evaluation . . . . .</b>	<b>43</b>
3.1	Dialogue and Interaction in the MeMo Workbench . . . . .	43
3.2	MeMo Principles . . . . .	44
3.2.1	Exchange of Information at Concept-Level . . . . .	45
3.2.2	Models . . . . .	46
3.2.3	Rules . . . . .	54
3.3	Usability-Evaluation with MeMo Workbench . . . . .	55
3.3.1	Describing User Interface Properties . . . . .	56
3.3.2	Modelling System Behaviour . . . . .	57
3.3.3	Defining User Tasks . . . . .	58
3.3.4	Defining User Groups . . . . .	59
3.3.5	Parameter Selection and Simulation . . . . .	60
3.3.6	Analysing Simulation Results . . . . .	61
<b>4</b>	<b>Logging Framework for the Evaluation of Interactive Systems . . . . .</b>	<b>63</b>
4.1	Model Concept . . . . .	65
4.1.1	Dialogue Abstraction Levels . . . . .	66
4.1.2	Representing the Dialogue Structure . . . . .	67
4.1.3	Parameters Describing Multimodal Interaction . . . . .	68
4.2	Data Structure and Model Design . . . . .	71
4.3	Implementation of the Framework . . . . .	74
4.4	Application to Empirical and Simulation Experiments . . . . .	76
4.4.1	Empirical Study I . . . . .	78
4.4.2	Empirical Study II . . . . .	83
4.4.3	Simulation-Based Experiment . . . . .	85
4.5	Discussion . . . . .	88
4.5.1	Review of Research Issues . . . . .	88
4.5.2	Completeness of PALADIN . . . . .	89
4.5.3	Application to Usability Studies . . . . .	92
4.5.4	Limitations in Automatic Logging of Interactions Parameters . . . . .	92
4.6	Chapter Summary . . . . .	93
<b>5</b>	<b>User Behaviour Model for the Evaluation of Interactive Systems . . . . .</b>	<b>95</b>
5.1	Method . . . . .	95
5.1.1	Empirical Study . . . . .	96
5.2	Results of the User Study . . . . .	105
5.2.1	Qualitative Results . . . . .	106
5.2.2	Quantitative Results . . . . .	112

- 5.3 Simulation . . . . . 114
  - 5.3.1 Technical-Sociological User Behaviour Model . . . . . 115
  - 5.3.2 Simulation Configuration . . . . . 116
  - 5.3.3 Task and User Group Definition . . . . . 117
  - 5.3.4 Simulation Results . . . . . 118
- 5.4 Discussion . . . . . 120
  - 5.4.1 Technical-Sociological Approach . . . . . 120
  - 5.4.2 Simulation and Simulation Results . . . . . 121
- 5.5 Chapter Summary . . . . . 121
- 6 Evaluation of User Behaviour Simulations . . . . . 123**
  - 6.1 Introduction . . . . . 123
  - 6.2 Mathematical Fundamentals . . . . . 126
    - 6.2.1 Frequency, Probability and Rank . . . . . 126
    - 6.2.2 Notation in Terms of Vectors . . . . . 127
    - 6.2.3 Comparing Frequency and Probability Distributions . . . . . 128
    - 6.2.4 Comparing Ranked Data . . . . . 131
  - 6.3 *n*-gram and *n*-gram Model . . . . . 135
    - 6.3.1 Comparing *n*-gram Models . . . . . 136
    - 6.3.2 Smoothing of *n*-gram Models . . . . . 138
  - 6.4 Method and Material . . . . . 140
    - 6.4.1 Dialogue Corpora . . . . . 140
    - 6.4.2 Dialogue Act Set and Interaction Parameters . . . . . 147
    - 6.4.3 Distance Measures . . . . . 148
    - 6.4.4 *n*-gram Model Creation . . . . . 148
    - 6.4.5 Distance Measure Evaluation . . . . . 154
    - 6.4.6 DiCoDis Framework . . . . . 157
    - 6.4.7 Application of the Most Appropriate Distance Measure . . . . . 157
  - 6.5 Experimental Results . . . . . 158
    - 6.5.1 Classifier Performance . . . . . 158
    - 6.5.2 Distance Between Corpora . . . . . 161
  - 6.6 Discussion . . . . . 164
    - 6.6.1 Method . . . . . 164
    - 6.6.2 Compliance with Evaluation Criteria . . . . . 167
    - 6.6.3 Requirements and Area of Application . . . . . 168
  - 6.7 Chapter Summary . . . . . 169
- 7 Conclusions . . . . . 171**
  - 7.1 Parameter-Based Description of Multimodal Interaction . . . . . 171
  - 7.2 Integration of Interactivity Patterns into User Behaviour Simulation . . . . . 173
  - 7.3 Evaluation of User Behaviour Simulation . . . . . 175
  - 7.4 Overall Conclusions . . . . . 176

**Appendix A: Summary of Guidelines for Multimodal Interaction**  
Description Languages . . . . . 179

**Appendix B: Participants in the TeSo Experiment.** . . . . . 181

**Appendix C: Material for TeSo Experiment.** . . . . . 183

**Appendix D: PALADIN Parameter Set.** . . . . . 201

**Appendix E: Screenshots of PALADIN Study II** . . . . . 213

**Appendix F: Translations for PALADIN Related Studies** . . . . . 215

**Appendix G: Equivalence of Mean and Symmetric Kullback-Leibler**  
Divergence. . . . . 217

**Appendix H: Distance Measures Evaluation and Application**  
Results . . . . . 219

**Appendix I: Additional Information on Dialogue Copra Addressed in**  
Chapter 6 . . . . . 223

**References** . . . . . 227