

Evaluation of Text and Speech Systems

Edited by

Laila Dybkjær

Prolog Development Center A/S

Brøndby

Denmark

Holmer Hensen

DFKI Language Technology Lab

Berlin

Germany

Wolfgang Minker

Institute for Information Technology

University of Ulm

Germany



Contents

Preface	ix
Contributing Authors	xi
Introduction	xv
<i>Laila Dybkjær, Holmer Hemsén and Wolfgang Minker</i>	
1	
Speech and Speaker Recognition Evaluation	1
<i>Sadaoki Furui</i>	
1. Introduction	1
2. Principles of Speech Recognition	2
3. Categories of Speech Recognition Tasks	3
4. Evaluation of Speech Recognition Systems	6
5. Principles of Speaker Recognition	12
6. Categories of Speaker Recognition Tasks	13
7. Normalization and Adaptation Techniques	15
8. Evaluation of Speaker Recognition Systems	17
9. Factors Affecting the Performance and Evaluation Paradigm Design for Speech and Speaker Recognition Systems	22
10. System-Level Evaluation of Speech and Speaker Recognition	23
11. Conclusion	24
References	24
2	
Evaluation of Speech Synthesis	29
<i>Nick Campbell</i>	
1. Introduction	29
2. Components of Computer Speech	30
3. Evaluation Methodologies	36
4. Organised Evaluations and Assessment	41
5. Speaking to (and on Behalf of) People	45
6. Conclusion	45
References	48

3

Modelling and Evaluating Verbal and Non-Verbal Communication
in Talking Animated Interface Agents 65*Björn Granström and David House*

1. Introduction 65
 2. KTH Parametric Multimodal Speech Synthesis 67
 3. Data Collection and Data-Driven Visual Synthesis 69
 4. Evaluating Intelligibility and Information Presentation 71
 5. Evaluating Visual Cues for Prominence 77
 6. Evaluating Prosody and Interaction 80
 7. Evaluating Visual Cues to Sentence Mode 84
 8. Evaluation of Agent Expressiveness and Attitude 85
 9. Agent and System Evaluation Studies 87
 10. Future Challenges in Modelling and Evaluation 91
- References 92

4

Evaluating Part-of-Speech Tagging and Parsing 99

Patrick Paroubek

1. POS Tagging 99
 2. Parsing 102
 3. Evaluation and Natural Language Processing 105
 4. POS Tagging Evaluation Methodology 110
 5. Methodology and Evaluation Measures for Parsing 114
 6. Conclusion 117
- References 118

5

General Principles of User-Oriented Evaluation 125

Margaret King

1. A Historical Note 126
 2. What is User-Oriented Evaluation? 128
 3. A First Principle: Quality is Decided by Users 129
 4. A Second Principle: Users do not Have the Same Needs 130
 5. A Third Principle: Quality can be Characterized 135
 6. A Fourth Principle: Quality can be Measured 148
 7. Combining the Particular and the General: The Ideal 153
 8. Conclusion 154
- References 156

6

An Overview of Evaluation Methods in TREC Ad Hoc Information
Retrieval and TREC Question Answering 163*Simone Teufel*

1. Introduction 163
 2. Evaluation Criteria 169
 3. Evaluation Metrics 172
 4. Real-World Performance 182
 5. Conclusion 183
- References 185

<i>Contents</i>	vii
7	
Spoken Dialogue Systems Evaluation	187
<i>Niels Ole Bernsen, Laila Dybkjær and Wolfgang Minker</i>	
1. Introduction	187
2. Evaluation Methods and Criteria	188
3. Evaluation of the NICE Hans Christian Andersen Prototype	193
4. Evaluation of the SENECA Prototype	207
5. Conclusion	216
References	218
8	
Linguistic Resources, Development, and Evaluation of Text and Speech Systems	221
<i>Christopher Cieri</i>	
1. Introduction	222
2. The Linguistic Resource Landscape	222
3. Background on Linguistic Data and Annotation	234
4. Data Planning for Technology Development and Evaluation	237
5. Finding Resources	240
6. Building Resources	242
7. Conclusion	259
References	260
9	
Towards International Standards for Language Resources	263
<i>Nancy Ide and Laurent Romary</i>	
1. Introduction	263
2. Background	265
3. The Linguistic Annotation Framework	268
4. Putting it All Together	278
5. Conclusion	282
References	283
Index	285