Contents

List of Contributors xiii

Preface xv

1 An Overview 1
Maxine Eskénazi

1.1 Origins of Crowdsourcing 2
1.2 Operational Definition of Crowdsourcing 3
1.3 Functional Definition of Crowdsourcing 3
1.4 Some Issues 4
1.5 Some Terminology 6
1.6 Acknowledgments 6
References 6

2 The Basics 8
Maxine Eskénazi

2.1 An Overview of the Literature on Crowdsourcing for Speech Processing 8
2.1.1 Evolution of the Use of Crowdsourcing for Speech 9
2.1.2 Geographic Locations of Crowdsourcing for Speech 10
2.1.3 Specific Areas of Research 12
2.2 Alternative Solutions 14
2.3 Some Ready-Made Platforms for Crowdsourcing 15
2.4 Making Task Creation Easier 17
2.5 Getting Down to Brass Tacks 17
2.5.1 Hearing and Being Heard over the Web 18
2.5.2 Prequalification 20
2.5.3 Native Language of the Workers 21
2.5.4 Payment 22
2.5.5 Choice of Platform in the Literature 25
2.5.6 The Complexity of the Task 27
2.6 Quality Control 29

	2.6.1 Was That Worker a Bot?	29
	2.6.2 Quality Control in the Literature	29
2.7	Judging the Quality of the Literature	32
2.8	Some Quick Tips	33
2.9	Acknowledgments	33
	References	33
	Further reading	35
3	**Collecting Speech from Crowds** *Ian McGraw*	**37**
3.1	A Short History of Speech Collection	38
	3.1.1 Speech Corpora	38
	3.1.2 Spoken Language Systems	40
	3.1.3 User-Configured Recording Environments	41
3.2	Technology for Web-Based Audio Collection	43
	3.2.1 Silverlight	44
	3.2.2 Java	45
	3.2.3 Flash	46
	3.2.4 HTML and JavaScript	48
3.3	Example: WAMI Recorder	49
	3.3.1 The JavaScript API	49
	3.3.2 Audio Formats	51
3.4	Example: The WAMI Server	52
	3.4.1 PHP Script	52
	3.4.2 Google App Engine	54
	3.4.3 Server Configuration Details	57
3.5	Example: Speech Collection on Amazon Mechanical Turk	59
	3.5.1 Server Setup	60
	3.5.2 Deploying to Amazon Mechanical Turk	61
	3.5.3 The Command-Line Interface	64
3.6	Using the Platform Purely for Payment	65
3.7	Advanced Methods of Crowdsourced Audio Collection	67
	3.7.1 Collecting Dialog Interactions	67
	3.7.2 Human Computation	68
3.8	Summary	69
3.9	Acknowledgments	69
	References	70
4	**Crowdsourcing for Speech Transcription** *Gabriel Parent*	**72**
4.1	Introduction	72
	4.1.1 Terminology	72
4.2	Transcribing Speech	73
	4.2.1 The Need for Speech Transcription	74
	4.2.2 Quantifying Speech Transcription	75

	4.2.3	*Brief History*	78
	4.2.4	*Is Crowdsourcing Well Suited to My Needs?*	79
4.3	Preparing the Data		80
	4.3.1	*Preparing the Audio Clips*	80
	4.3.2	*Preprocessing the Data with a Speech Recognizer*	81
	4.3.3	*Creating a Gold-Standard Dataset*	82
4.4	Setting Up the Task		83
	4.4.1	*Creating Your Task with the Platform Template Editor*	83
	4.4.2	*Creating Your Task on Your Own Server*	85
	4.4.3	*Instruction Design*	87
	4.4.4	*Know the Workers*	89
	4.4.5	*Game Interface*	91
4.5	Submitting the Open Call		91
	4.5.1	*Payment*	92
	4.5.2	*Number of Distinct Judgments*	93
4.6	Quality Control		95
	4.6.1	*Normalization*	95
	4.6.2	*Unsupervised Filters*	96
	4.6.3	*Supervised Filters*	99
	4.6.4	*Aggregation Techniques*	100
	4.6.5	*Quality Control Using Multiple Passes*	101
4.7	Conclusion		102
4.8	Acknowledgments		103
	References		103
5	**How to Control and Utilize Crowd-Collected Speech**		**106**
	Ian McGraw and Joseph Polifroni		
5.1	Read Speech		107
	5.1.1	*Collection Procedure*	107
	5.1.2	*Corpus Overview*	108
5.2	Multimodal Dialog Interactions		111
	5.2.1	*System Design*	111
	5.2.2	*Scenario Creation*	111
	5.2.3	*Data Collection*	112
	5.2.4	*Data Transcription*	115
	5.2.5	*Data Analysis*	118
5.3	Games for Speech Collection		120
5.4	Quizlet		121
5.5	Voice Race		123
	5.5.1	*Self-Transcribed Data*	124
	5.5.2	*Simplified Crowdsourced Transcription*	124
	5.5.3	*Data Analysis*	125
	5.5.4	*Human Transcription*	126
	5.5.5	*Automatic Transcription*	127
	5.5.6	*Self-Supervised Acoustic Model Adaptation*	127

5.6 Voice Scatter 129
5.6.1 Corpus Overview 130
5.6.2 Crowdsourced Transcription 131
5.6.3 Filtering for Accurate Hypotheses 132
5.6.4 Self-Supervised Acoustic Model Adaptation 133
5.7 Summary 135
5.8 Acknowledgments 135
References 136

6 Crowdsourcing in Speech Perception **137**
Martin Cooke, Jon Barker, and Maria Luisa Garcia Lecumberri
6.1 Introduction 137
6.2 Previous Use of Crowdsourcing in Speech and Hearing 138
6.3 Challenges 140
6.3.1 Control of the Environment 140
6.3.2 Participants 141
6.3.3 Stimuli 144
6.4 Tasks 145
6.4.1 Speech Intelligibility, Quality and Naturalness 145
6.4.2 Accent Evaluation 146
6.4.3 Perceptual Salience and Listener Acuity 147
6.4.4 Phonological Systems 147
6.5 BigListen: A Case Study in the Use of Crowdsourcing to Identify Words in Noise 149
6.5.1 The Problem 149
6.5.2 Speech and Noise Tokens 150
6.5.3 The Client-Side Experience 150
6.5.4 Technical Architecture 151
6.5.5 Respondents 153
6.5.6 Analysis of Responses 158
6.5.7 Lessons from the BigListen Crowdsourcing Test 166
6.6 Issues for Further Exploration 167
6.7 Conclusions 169
References 169

7 Crowdsourced Assessment of Speech Synthesis **173**
Sabine Buchholz, Javier Latorre, and Kayoko Yanagisawa
7.1 Introduction 173
7.2 Human Assessment of TTS 174
7.3 Crowdsourcing for TTS: What Worked and What Did Not 177
7.3.1 Related Work: Crowdsourced Listening Tests 177
7.3.2 Problem and Solutions: Audio on the Web 178
7.3.3 Problem and Solution: Test of Significance 180
7.3.4 What Assessment Types Worked 183
7.3.5 What Did Not Work 186

7.3.6 *Problem and Solutions: Recruiting Native Speakers of Various Languages* 190
7.3.7 *Conclusion* 193
7.4 Related Work: Detecting and Preventing Spamming 193
7.5 Our Experiences: Detecting and Preventing Spamming 195
7.5.1 *Optional Playback Interface* 196
7.5.2 *Investigating the Metrics Further: Mandatory Playback Interface* 201
7.5.3 *The Prosecutor's Fallacy* 210
7.6 Conclusions and Discussion 212
References 214

8 Crowdsourcing for Spoken Dialog System Evaluation **217**
Zhaojun Yang, Gina-Anne Levow, and Helen Meng

8.1 Introduction 217
8.2 Prior Work on Crowdsourcing: Dialog and Speech Assessment 220
8.2.1 *Prior Work on Crowdsourcing for Dialog Systems* 220
8.2.2 *Prior Work on Crowdsourcing for Speech Assessment* 220
8.3 Prior Work in SDS Evaluation 221
8.3.1 *Subjective User Judgments* 221
8.3.2 *Interaction Metrics* 222
8.3.3 *PARADISE Framework* 223
8.3.4 *Alternative Approach to Crowdsourcing for SDS Evaluation* 224
8.4 Experimental Corpus and Automatic Dialog Classification 225
8.5 Collecting User Judgments on Spoken Dialogs with Crowdsourcing 226
8.5.1 *Tasks for Dialog Evaluation* 227
8.5.2 *Tasks for Interannotator Agreement* 229
8.5.3 *Approval of Ratings* 229
8.6 Collected Data and Analysis 230
8.6.1 *Approval Rates and Comments from Workers* 230
8.6.2 *Consistency between Automatic Dialog Classification and Manual Ratings* 231
8.6.3 *Interannotator Agreement among Workers* 233
8.6.4 *Interannotator Agreement on the Let's Go! System* 235
8.6.5 *Consistency between Expert and Nonexpert Annotations* 236
8.7 Conclusions and Future Work 238
8.8 Acknowledgments 238
References 239

9 Interfaces for Crowdsourcing Platforms **241**
Christoph Draxler

9.1 Introduction 241
9.2 Technology 242
9.2.1 *TinyTask Web Page* 242
9.2.2 *World Wide Web* 242
9.2.3 *Hypertext Transfer Protocol* 243

	9.2.4	*Hypertext Markup Language*	244
	9.2.5	*Cascading Style Sheets*	246
	9.2.6	*JavaScript*	246
	9.2.7	*JavaScript Object Notation*	248
	9.2.8	*Extensible Markup Language*	248
	9.2.9	*Asynchronous JavaScript and XML*	249
	9.2.10	*Flash*	250
	9.2.11	*SOAP and REST*	251
	9.2.12	*Section Summary*	252
9.3	Crowdsourcing Platforms		253
	9.3.1	*Crowdsourcing Platform Workflow*	253
	9.3.2	*Amazon Mechanical Turk*	256
	9.3.3	*CrowdFlower*	259
	9.3.4	*Clickworker*	259
	9.3.5	*WikiSpeech*	260
9.4	Interfaces to Crowdsourcing Platforms		261
	9.4.1	*Implementing Tasks Using a GUI on the CrowdFlower Platform*	262
	9.4.2	*Implementing Tasks Using the Command-Line Interface in MTurk*	264
	9.4.3	*Implementing a Task Using a RESTful Web Service in Clickworker*	270
	9.4.4	*Defining Tasks via Configuration Files in WikiSpeech*	270
9.5	Summary		278
	References		278
10	**Crowdsourcing for Industrial Spoken Dialog Systems**		**280**
	David Suendermann and Roberto Pieraccini		
10.1	Introduction		280
	10.1.1	*Industry's Willful Ignorance*	280
	10.1.2	*Crowdsourcing in Industrial Speech Applications*	281
	10.1.3	*Public versus Private Crowd*	282
10.2	Architecture		283
10.3	Transcription		287
10.4	Semantic Annotation		290
10.5	Subjective Evaluation of Spoken Dialog Systems		296
10.6	Conclusion		300
	References		300
11	**Economic and Ethical Background of Crowdsourcing for Speech**		**303**
	Gilles Adda, Joseph J. Mariani, Laurent Besacier, and Hadrien Gelas		
11.1	Introduction		303
11.2	The Crowdsourcing Fauna		304
	11.2.1	*The Crowdsourcing Services Landscape*	304
	11.2.2	*Who Are the Workers?*	306
	11.2.3	*Ethics and Economics in Crowdsourcing: How to Proceed?*	307
11.3	Economic and Ethical Issues		307
	11.3.1	*What Are the Problems for the Workers?*	309

	11.3.2	*Crowdsourcing and Labor Laws*	310
	11.3.3	*Which Economic Model Is Sustainable for Crowdsourcing?*	314
11.4	Under-Resourced Languages: A Case Study		316
	11.4.1	*Under-Resourced Languages Definition and Issues*	317
	11.4.2	*Collecting Annotated Speech for African Languages Using Crowdsourcing*	317
	11.4.3	*Experiment Description*	317
	11.4.4	*Results*	318
	11.4.5	*Discussion and Lessons Learned*	321
11.5	Toward Ethically Produced Language Resources		322
	11.5.1	*Defining a Fair Compensation for Work Done*	323
	11.5.2	*Impact of Crowdsourcing on the Ecology of Linguistic Resources*	326
	11.5.3	*Defining an Ethical Framework: Some Solutions*	326
11.6	Conclusion		330
	Disclaimer		331
	References		331

Index **335**