

**Signals
and
Communication
Technology**

**D. Gardner-Bonneau
H. E. Blanchard (Eds.)**

Human Factors and Voice Interactive Systems

Second Edition

 **Springer**

CONTENTS

1. IVR Usability Engineering Using Guidelines and Analyses of End-To-End Calls.....	1
<i>Bernhard Suhm</i>	
1. IVR Design Principles and Guidelines	2
1.1 A Taxonomy of Limitations of Speech User Interfaces	3
1.1.1 Limitations of Speech Recognition	4
1.1.2 Limitations of Spoken Language	7
1.1.3 Human Cognition	9
1.2 Towards Best Practices for IVR Design	10
1.2.1 A Database for Speech User Interface Design Knowledge	10
1.2.2 Compiling Guidelines for IVR Design	11
1.2.3 Applying IVR Design Guidelines in Practice	13
1.3 Best Practices for IVR Design?	18
2. Data-Driven IVR Usability Engineering Based on End-To-End Calls.....	19
2.1 The Flaws of Standard IVR Reports	20
2.2 Capturing End-to-End Data from Calls	20
2.3 Evaluating IVR Usability based on End-to-End Calls.....	23
2.3.1 Call-reason Distribution	23
2.3.2 Diagnosing IVR Usability using Caller-Path Diagrams	24
2.3.3 IVR Usability Analysis using Call-Reason Distribution and Caller-Path Diagrams	27
2.4 Evaluating IVR Cost-effectiveness.....	29
2.4.1 Defining Total IVR Benefit	30
2.4.2 Measuring Total IVR Benefit.....	31
2.4.3 Estimating Improvement Potential.....	34
2.4.4 Building the Business Case for IVR Redesign.....	35
3. Summary and Conclusions.....	38
Acknowledgements.....	39
References.....	39

2. User Interface Design for Natural Language Systems: From Research to Reality.....	43
<i>Susan J. Boyce</i>	
1. Introduction.....	43
1.1 What is Natural Language?	43
1.1.1 Natural Language for Call Routing.....	44
1.1.2 Natural Language for Form Filling.....	45
1.1.3 The Pros and Cons of Natural Language Interfaces	45
1.2 What Are the Steps to Building a Natural Language Application?.....	46
1.2.1 Data Collection	46
1.2.2 Annotation Guide Development	47
1.2.3 Call Flow Development and Annotation.....	48
1.2.4 Application Code and Grammar/NL Development.....	49
1.2.5 Testing NL Applications	49
1.2.6 Post-Deployment Tuning	49
1.3 When Does it Make Sense to use Natural Language?	50
1.3.1 Distribution of Calls.....	50
1.3.2 Characteristics of the Caller Population.....	51
1.3.3 Evidence Obtained from Data with Existing Application.....	53
1.3.4 Ease of Getting to an Agent	53
1.3.5 Live Caller Environment Versus IVR: What is Being Replaced?.....	53
1.4 The Call Routing Task.....	54
1.5 Design Process.....	54
1.6 Analysis of Human-to-Human Dialogues.....	55
2. Anthropomorphism and User Expectations	55
2.1 Anthropomorphism Experiment.....	56
3. Issues for Natural Dialogue Design	60
3.1 Initial Greeting.....	60
3.2 Confirmations	60
3.3 Disambiguating an Utterance	61
3.4 Reprompts.....	61
3.5 Turn-taking.....	62
3.6 When to Bail Out.....	62
4. Establishing User Expectations in the Initial Greeting	62
4.1 Initial Greeting Experiment	63
5. Identifying Recognition Errors Through Confirmations.....	66
5.1 Confirming Digit Strings in Spoken Dialogue Systems.....	67

5.2	Confirmation of Topic in a Spoken Natural Dialogue System	69
6.	Repairing Recognition Errors With Reprompts	72
6.1	Reprompt Experiment	73
7.	Turn-Taking in Human-Machine Dialogues	76
7.1	Caller Tolerance of System Delay	77
8.	Summary	79
	References	79
3.	Linguistics and Psycholinguistics in IVR Design	81
	<i>Osamuyimen T. Stewart and Harry E. Blanchard</i>	
1.	Introduction	82
1.1	Speech Sounds	82
1.2	Grammar	83
1.2.1	Words	84
1.2.2	Sentences	84
1.2.3	Meaning	85
2.	ASR Grammars and Language Understanding	86
2.1	Morphology	87
2.2	Syntax	88
2.3	Semantics	93
2.3.1	Synonyms	93
2.3.2	Polysemy	94
2.4	Putting it All Together	94
2.5	ASR Grammars	95
2.6	Natural Language Understanding Models	97
2.6.1	The Semantic Taxonomy	98
2.6.2	Establishing Predicates	100
3.	Dialog Design	102
3.1	Putting it All Together	105
3.1.1	Scenario 1	106
3.1.2	Scenario 2	107
4.	Consequences of Structural Simplification	108
4.1	Semantic Specificity	111
4.2	Syntactic Specificity	112
	Conclusion	113
	References	113
4.	Designing the Voice User Interface for Automated Directory Assistance	117
	<i>Amir M. Mané and Esther Levin</i>	
1.	The Business of DA	117
1.1	The Introduction of Automation	118

1.2	Early Attempts to Use Speech Recognition.....	119
2.	Issues in the Design of VUI for DA	121
2.1	Addressing Database Inadequacies.....	122
2.1.1	The Solution: Automated Data Cleaning	123
2.2	Pronunciation of Names	123
2.3	The First Question	124
2.4	Finding the Locality.....	124
2.5	Confirming the Locality	125
2.6	Determining the Listing Type.....	126
2.7	Handling Business Requests.....	127
2.7.1	Issues in Grammar Design for Business Listing Automation	127
2.7.2	Business Listings Disambiguation	130
2.8	Handling Residential Listings.....	131
2.9	General Dialogue Design Issues	133
3.	Final Thoughts	134
	References.....	134
5.	Spoken Language Interfaces for Embedded Applications	135
	<i>Dragos Burileanu</i>	
1.	Introduction.....	135
2.	Spoken Language Interfaces Development.....	137
2.1	Overview. Current Trends	137
2.2	Embedded Speech Applications	139
3.	Embedded Speech Technologies.....	141
3.1	Technical Constraints and Implementation Methods	141
3.2	Embedded Speech Recognition	143
3.3	Embedded Speech Synthesis	149
4.	A Case Study: An Embedded TTS System Implementation.....	153
4.1	A Simplified TTS System Architecture.....	153
4.2	Implementation Issues	155
5.	The Future of Embedded Speech Interfaces	158
	References.....	160
6.	Speech Generation in Mobile Phones.....	163
	<i>Géza Németh, Géza Kiss, Csaba Zainkó, Gábor Olaszy, and Bálint Tóth</i>	
1.	Introduction.....	163
2.	Speaking Telephone? What is it Good for?	165
3.	Speech Generation Technologies in Mobile Phones.....	166
3.1	Synthesis Technologies	167
3.1.1	Limited Vocabulary Concatenation	167
3.1.2	Unlimited Text Reading - Text-To-Speech.....	168
3.2	Topic-Related Text Preprocessing.....	170

3.2.1	Exceptions Vocabulary	171
3.2.2	Complex Text Transformation.....	171
3.2.3	Language Identification	174
4.	How to Port Speech Synthesis on a Phone Platform.....	178
5.	Limitations and Possibilities Offered by Phone Resources	181
6.	Implementations.....	183
6.1	The Mobile Phone as a Speaking Aid.....	183
6.2	An SMS-Reading Mobile Phone Application	186
	Acknowledgements.....	190
	References.....	190
7.	Voice Messaging User Interface.....	193
	<i>Harry E. Blanchard and Steven H. Lewis</i>	
1.	Introduction.....	193
2.	The Touch-Tone Voice Mail user Interface.....	196
2.1	Common Elements of Touch-tone Transactions	197
2.1.1	Prompts	197
2.1.2	Interruptibility	198
2.1.3	Time-outs and Reprompts.....	199
2.1.4	Feedback	200
2.1.5	Feedback to Errors	200
2.1.6	Menu Length.....	200
2.1.7	Mapping of Keys to Options	201
2.1.8	Global Commands.....	201
2.1.9	Use of the “#” and “*” Keys	202
2.1.10	Unprompted Options.....	202
2.1.11	Voice and Personality	203
2.2	Call Answering.....	203
2.2.1	Call Answering Greetings.....	206
2.3	The Subscriber Interface.....	206
2.4	Retrieving and Manipulating Messages.....	206
2.5	Sending Messages.....	209
2.6	Voice Messaging User Interface Standards	211
2.7	Alternative Approaches to Traditional Touch-tone Design	214
3.	Automatic Speech Recognition and Voice Mail	215
4.	Unified Messaging and Multimedia Mail	219
4.1	Fax Messaging	220
4.2	Viewing Voice Mail	221
4.3	Listening to E-mail	223
4.4	Putting it All Together	224
4.5	Mixed Media.....	225
	References.....	226

8. Silence Locations and Durations in Dialog Management	231
<i>Matthew Yuschik</i>	
1. Introduction.....	231
2. Prompts and Responses in Dialog Management	233
2.1 Dialog Management.....	233
2.2 Word Selection	234
2.3 Word Lists	234
2.4 Turn-Taking Cues.....	236
3. Time as an Independent Variable – Dialog Model	236
3.1 Definition of Terms	237
3.2 Examples of Usage	238
4. User Behavior	238
4.1 Transactional Analysis.....	238
4.2 Verbal Communication.....	239
4.3 Directed Dialogs	239
5. Measurements	240
5.1 Barge-In	241
6. Usability Testing and Results.....	242
6.1 Test Results - United States (early prototype)	244
6.2 Test Results – United States (tuned, early prototype).....	245
6.3 Test Results - United Kingdom.....	246
6.4 Test Results – Italy	247
6.5 Test Results – Denmark.....	249
7. Observations and Interpretations.....	250
7.1 Lateral Results	250
7.2 Learning – Longitudinal Results	251
Conclusions.....	252
Acknowledgement	252
References.....	252

9. Using Natural Dialogs as the Basis for Speech Interface Design ... 255

Nicole Yankelovich

1. Introduction.....	256
1.1 Motivation	256
1.2 Natural Dialog Studies.....	257
2. Natural Dialog Case Studies	258
2.1 Study #1: SpeechActs Calendar (speech-only, telephone-based)	259
2.1.1 Purpose of Application.....	259
2.1.2 Study Design.....	260
2.1.3 Software Design.....	262
2.1.4 Lessons Learned.....	264

2.2	Study #2: Office Monitor (speech-only, microphone-based)	264
2.2.1	Purpose of Application.....	264
2.2.2	Study Design.....	265
2.2.3	Software Design.....	267
2.2.4	Lessons Learned.....	269
2.3	Study #3: Automated Customer Service Representative (speech input, speech/graphical output, telephone-based)	269
2.3.1	Purpose of Application.....	269
2.3.2	Study Design.....	269
2.3.3	Software Design.....	275
2.3.4	Lessons Learned.....	278
2.4	Study #4: Multimodal Drawing (speech/mouse/keyboard input, speech/graphical output, microphone-based)	278
2.4.1	Purpose of Application.....	278
2.4.2	Study Design.....	279
2.4.3	Software Design.....	283
2.4.4	Lessons Learned.....	286
3.	Discussion.....	286
3.1	Refining Application Requirements and Functionality.....	286
3.2	Collecting Appropriate Vocabulary.....	287
3.3	Determining Commonly used Grammatical Constructs.....	287
3.4	Discovering Effective Interaction Patterns.....	287
3.5	Helping with Prompt and Feedback Design	288
3.6	Getting a Feeling for the Tone of the Conversations.....	288
	Conclusion	289
	Acknowledgements.....	289
	References.....	290

10. Telematics: Artificial Passenger and Beyond..... 291

Dimitri Kanevsky

1.	Introduction.....	291
2.	A Brief Overview of IBM Voice Technologies	292
2.1	Conversational Interactivity for Telematics	293
2.2	System Architecture.....	295
2.3	Embedded Speech Recognition	297
2.4	Distributed Speech Recognition	299
3.	Evaluating/Predicting the Consequences of Misrecognitions.....	300

4.	Improving Voice and State Recognition Performance – Network Data Collection, Learning by Example, Adaptation of Language and Acoustic Models for Similar users	303
5.	Artificial Passenger	308
6.	User Modeling Aspects	315
6.1	User Model	316
6.2	The Adaptive Modeling Process	317
6.3	The Control Process	318
6.4	Discussion about Time-Lagged Observables and Indicators in a History	319
7.	Gesture-Based Command Interface	320
8.	Summary	322
	Acknowledgements	323
	References	323

11. A Language to Write Letter-To-Sound Rules for English and French..... 327

Michel Divay

1.	Introduction	327
2.	The Historic Evolution of English and French	329
3.	The Complexity of the Conversion for English and French	329
4.	Rule Formalism	334
5.	Examples of Rules for English	340
6.	Examples of Rules for French	345
	Conclusions	353
	References	354
	Appendices for French	356
	Appendices for English	359

12. Virtual Sentences of Spontaneous Speech: Boundary Effects of Syntactic-Semantic-Prosodic Properties..... 361

Mária Gósy and Magdolna Kovács

1.	Introduction	361
2.	Method and Material	364
2.1	Subjects	364
2.2	Speech Material	364
2.3	Procedure	365
3.	Results	366
3.1	Identification of Virtual Sentences in the Normal and Filtered Speech Samples	366
3.2	Pauses of the Speech Sample	368
3.3	Pause Perception	370

3.4	F0 Patterns	372
3.5	Comprehension of the Spontaneous Speech Sample	374
3.6	The Factor of Gender	375
	Conclusions	375
	Acknowledgements	377
	References	377

13. Text-to-Speech Formant Synthesis For French 381 *Michel Divay and Ed Bruckert*

1.	Introduction	381
2.	Grapheme-to-Phoneme Conversion	382
2.1	Normalization: From Grapheme to Grapheme	382
2.2	From Grapheme to Phoneme	384
2.3	Exception Dictionary	385
3.	Prosody	385
3.1	Parsing the Text	385
3.2	Intonation	386
3.3	Phoneme Duration	391
4.	Acoustics for French Consonants and Vowels	398
4.1	Vowels	398
4.2	Fricatives (unvoiced: F, S, Ch; voiced: V, Z, J)	400
4.3	Plosives (unvoiced: P, T, K; voiced: B, D, G)	401
4.4	Nasals (M, N, Gn, Ng)	403
4.5	Liquids (L, R)	404
4.6	Semivowels (Y, W, Wu)	405
4.7.	Phoneme Transitions (coarticulation effects)	405
4.8	Frame Generation	409
4.9	Conclusions for Acoustics	409
5.	From Acoustics to Speech Signal	410
6.	Next Generation Formant Synthesis	412
7.	Singing	414
	Conclusions	414
	References	415

14. Accessibility and Speech Technology: Advancing Toward Universal Access 417 *John C. Thomas, Sara Basson, and Daryle Gardner-Bonneau*

1.	Universal Access vs. Assistive Technology	417
2.	Predicted Enhancements and Improvements to Underlying Technology	419
2.1	Social Network Analysis, Blogs, Wikis, and Social Computing	420
2.2	Intelligent Agents	421

2.3	Learning Objects.....	422
2.4	Cognitive Aids.....	423
2.5	Interface Flexibility and Intelligence.....	423
3.	Current Assistive Technology Applications Employing Speech Technology.....	423
3.1	Applications Employing Automatic Speech Recognition (ASR).....	424
3.2	Applications of Synthetic Speech.....	428
4.	Human-Computer Interaction: Design and Evaluation.....	430
5.	The Role of Technical Standards in Accessibility.....	433
5.1	Standards Related to Software and Information Technology User Interfaces.....	434
5.2	Speech Application Accessibility Standards.....	434
5.3	Accessibility Data and Accessibility Guidance for General Products.....	437
	Conclusions.....	439
	References.....	440

15. Synthesized Speech Used for the Evaluation of Children's Hearing and Speech Perception 443

Mária Gósy

1.	Introduction.....	443
2.	The Background Theory.....	444
3.	The Production of the Synthesized Word Material.....	447
4.	Pre-Experiments for the Application of Synthesized Words for Hearing Screening.....	449
5.	Results.....	450
5.1	Clinical Tests.....	450
5.2	Screening Procedure.....	453
5.3	Evaluation of Acoustic-phonetic Perception.....	456
5.4	Children with Specific Needs.....	457
	Conclusions.....	458
	Acknowledgements.....	459
	References.....	459

Index 461