

Contents

1	User's guide	1
1.1	Background	1
1.1.1	EAGLES objectives	1
1.1.2	EAGLES organisational structure	2
1.1.3	EAGLES workplan	3
1.2	Spoken Language systems, standards and resources	4
1.2.1	Spoken Language systems	4
1.2.2	Standards and resources for Spoken Language systems	5
1.3	The EAGLES Spoken Language Working Group (WG5)	9
1.3.1	Subgroups of the EAGLES Spoken Language Working Group	11
1.3.2	Relationships with the other EAGLES Working Groups	11
1.3.3	Workshops	11
1.3.4	Production of the handbook	13
1.3.5	Consultation with the R&D Community	13
1.4	Overview of the handbook	14
1.4.1	Intended readership	14
1.4.2	Scope	15
1.4.3	The main chapters of the handbook	21
1.5	The current state of play	25
1.6	Possible future actions	25
1.6.1	Revision and completion of existing documentation	25
1.6.2	Extended survey of existing practice	25
1.6.3	Extension of language base	26
1.6.4	Terminology	26
1.6.5	Move to prescriptive recommendations	26
1.6.6	Publication and dissemination	26
1.6.7	Coordination with other bodies	26
1.7	Contact points	27
1.8	Acknowledgements	28
	Part I: Spoken language system and corpus design	29
2	System design	30
2.1	Introduction	30
2.1.1	System capability profile versus application requirement profile	32
2.1.2	Technical features versus system capabilities	33
2.1.3	System in operation versus laboratory prototype	34

2.2	Speech recognition systems	34
2.2.1	Speaker dependency	35
2.2.2	Speaker-adapted systems	36
2.2.3	Speaking aspects	37
2.2.4	Vocabulary aspects	38
2.2.5	Branching factor (perplexity factor)	41
2.2.6	Rejection mode	41
2.2.7	Application vocabulary and confusion matrix	42
2.2.8	Speaker selection	43
2.2.9	Language modelling	43
2.2.10	Channel adaptation/Environment adaptation	43
2.2.11	Task/Application adaptation	44
2.2.12	Speech recognition and application interfaces	44
2.2.13	Speech input and speech signal acquisition	44
2.2.14	Cut-through versus voice-stop (anticipation)	48
2.2.15	Error measure presentation	49
2.2.16	Error handling	49
2.2.17	Response time	50
2.3	Speaker verification/identification	51
2.3.1	Speaker verification versus speaker identification	51
2.3.2	Speaker characteristics verification or identification	51
2.3.3	Error measure	51
2.3.4	Training	53
2.3.5	Exploitation	54
2.4	Speech synthesis	56
2.4.1	Speech recording, storage, and playback	56
2.4.2	Canned speech	58
2.4.3	Text-to-speech synthesis	58
2.4.4	Multi-linguality	61
2.5	Interactive voice systems	61
2.5.1	Menu-driven systems	61
2.5.2	Sophisticated dialogue manager	62
2.5.3	Interaction with an information database	62
2.5.4	Error recovery strategy	62
2.5.5	Routing to an operator	63
2.5.6	DTMF, pulse detection, and voice recognition	63
2.6	System platform	63
2.6.1	Software aspects	63
2.6.2	Hardware aspects	64
2.6.3	Planning for expansion	67
2.7	System simulation and prototyping	67
2.8	Host-computer interfaces	67
2.9	Computer telephony integration	67

2.10	Multi-lingual aspects	68
2.11	System dimension configuration	68
2.12	Statistical tools	70
2.13	Conclusion	70
2.14	Recommendation: Requirement profiles	70
3	SL corpus design	79
3.1	Introduction	79
3.1.1	Spoken language corpus	79
3.1.2	About this chapter	80
3.2	Eight main differences between collections of written and spoken language data	81
3.2.1	Durability of text, volatility of speech	82
3.2.2	Different production times for text and speech	82
3.2.3	Correcting errors in the production of text and speech	82
3.2.4	Orthographic identity and phonetic variability of lexicalised units	83
3.2.5	Printable ASCII-strings and continuously sampled speech	84
3.2.6	Size differences between NL and SL data	85
3.2.7	The different legal status of written texts and spoken words	85
3.2.8	The different nature of categories and time functions	86
3.3	Applications of spoken language corpora	86
3.3.1	Speech corpora for research purposes	87
3.3.2	Speech corpora for technological applications	92
3.4	Specification of the linguistic content	99
3.4.1	Different types of speech data	99
3.4.2	Factorial experiments and corpus studies	106
3.5	Specification of number and type of speakers	107
3.5.1	Corpus size in terms of speakers	107
3.5.2	Speaker characteristics	110
3.6	Final comments	118
4	SL corpus collection	119
4.1	Introduction	119
4.2	Data collection dimensions	120
4.2.1	Visibility: open vs. secret	120
4.2.2	Clandestine microphone recordings	120
4.2.3	Visible recordings	121
4.2.4	Environment: Studio vs. on location	122
4.2.5	Interaction and control	125
4.2.6	Monitoring and validation	129

4.2.7	Multi-channel recording	130
4.3	Procedures	132
4.3.1	Equipment	132
4.3.2	Management	138
4.3.3	Wizard of Oz	143
4.3.4	Legal aspects	143
5	SL corpus representation	146
5.1	Introduction	146
5.1.1	The transcription of read speech versus the transcription of spontaneous speech	147
5.1.2	Levels and types of transcription	149
5.2	Concerning the segmentation and labelling of speech signals	150
5.2.1	Background	150
5.2.2	A caveat	151
5.2.3	Use of segmented speech databases	151
5.2.4	The CRIL conventions	151
5.2.5	Manual segmentation	152
5.2.6	Automatic and semi-automatic segmentation	153
5.2.7	Segmentation and labelling in the VERBMOBIL project	154
5.3	The levels and types of symbolic representations of speech	155
5.3.1	Recording script	155
5.3.2	Orthographic transcription	155
5.3.3	Morphological, syntactic, semantic and pragmatic representations	158
5.3.4	Citation-phonemic representation	158
5.3.5	Broad phonetic or phonotypic transcription	159
5.3.6	Narrow phonetic transcription	160
5.3.7	Acoustic-phonetic transcription	160
5.3.8	Physical transcription	161
5.3.9	Prosodic transcription	161
5.3.10	Non-linguistic and other phenomena	168
5.4	List of recommendations	170

Part II: Spoken language characterisation 175

6	Spoken language lexica	176
6.1	Introduction	176
6.1.1	Lexica for spoken language systems	176
6.1.2	Lexical information as properties of words	177
6.1.3	Recommendations on resources	183
6.2	What is a spoken language lexicon?	184
6.2.1	Basic features of a spoken language lexicon	184

6.2.2	Lexical databases and system lexica for spoken language	185
6.2.3	Spoken language and written language lexica	187
6.2.4	Basic lexicographic coverage criteria	189
6.2.5	The lexicon in spoken language recognition systems .	190
6.2.6	Recommendations on defining spoken language lexica	191
6.3	Types of lexical information in spoken language lexica	192
6.3.1	Lexicon models and lexical representation	192
6.3.2	A simple sign model for lexical properties	193
6.3.3	Lexical units	195
6.3.4	Lexical properties and lexical relations in spoken language	200
6.3.5	Recommendations on types of lexical information . . .	202
6.4	Lexical surface information	203
6.4.1	Orthographic information	203
6.4.2	Pronunciation information	204
6.4.3	Prosodic information	209
6.4.4	Recommendations on lexical surface information . . .	211
6.5	Morphological information	212
6.5.1	Types of morphological information	212
6.5.2	Applications of morphology	213
6.5.3	Recommendations on morphology	216
6.6	Grammatical information	216
6.6.1	Statistical language models	216
6.6.2	Sentence syntax information	217
6.6.3	Recommendations on grammatical information	218
6.7	Lexical content information	219
6.7.1	Lexical semantic information	219
6.7.2	Pragmatic information	219
6.7.3	Idiomatic information	220
6.7.4	Recommendations on semantic information	221
6.8	Lexicon structure	221
6.8.1	Spoken language lexicon formalisms	221
6.8.2	Lexicon architecture and lexical database structure . .	223
6.8.3	Lexicon architecture and the structure of lexical databases	225
6.8.4	Recommendations on lexicon structure	231
6.9	Lexical knowledge acquisition for spoken language	231
6.9.1	Stages in lexical knowledge acquisition	231
6.9.2	Types of knowledge source	232
6.9.3	Recommendations on lexicon construction	234
6.10	Outlook	235

7.1	Introduction	237
7.1.1	Goals of the chapter	237
7.1.2	Advice to the reader	238
7.2	System architecture for speech recognition	238
7.2.1	Bayes decision rule	238
7.2.2	Stochastic language modelling	240
7.2.3	Types of language models	243
7.2.4	Why does the trigram language model work?	244
7.3	Perplexity definition	245
7.3.1	Formal definition	245
7.3.2	Implications of the perplexity definition	246
7.3.3	Extensions of the perplexity definition	247
7.3.4	Recommendations	247
7.4	Language model smoothing: modelling unseen events	248
7.4.1	Problem formulation	248
7.4.2	Linear discounting and backing-off	249
7.4.3	Linear interpolation	252
7.4.4	Absolute discounting and backing-off	252
7.4.5	Conclusion	253
7.5	Multilevel smoothing for trigram models	253
7.5.1	Problem formulation	253
7.5.2	The full trigram model	254
7.5.3	Practical issues	255
7.5.4	Cache	256
7.5.5	Experimental results	257
7.5.6	Recommendations: m -gram language models	261
7.6	Refined language models	262
7.6.1	Equivalence histories and word classes	262
7.6.2	Grammar based language models	266
7.6.3	Recommendations: Refined language models	270
7.7	Language models and search	270
7.7.1	Sources of recognition errors	270
7.7.2	Search: Finding the single best sentence	271
7.7.3	Search: Word graphs	275
7.7.4	Recommendations: Language models and search	281
7.8	Final note: the mathematics of smoothing	281
7.8.1	Linear discounting and backing-off	281
7.8.2	Linear interpolation	284
7.8.3	Absolute discounting and backing-off	286
8	Physical characterisation and description	
8.1	Introduction	288
8.1.1	The communication chain	288
		289

8.1.2	Specifications for speech corpora production and speech assessment environments	290
8.1.3	Recommendations on recording strategies	291
8.2	Basic notations and terminology	291
8.2.1	Periodic and aperiodic motion	291
8.2.2	The dB scale	292
8.2.3	Phones and sones	293
8.2.4	Analog and digital speech signal representation	293
8.3	Talker / listener descriptors	295
8.3.1	General (demographic) descriptors	295
8.3.2	Task specific descriptors	296
8.4	Transducer characteristics	301
8.4.1	Microphones and speech recordings	302
8.4.2	Parallel recordings	317
8.5	Environment characteristics	321
8.5.1	Artificial vs. natural environment	321
8.5.2	Auditory environment	322
8.5.3	Visual environment	330
8.5.4	Miscellaneous environmental factors	331
8.6	Recording chain characteristics	331
8.6.1	Signals and systems	332
8.6.2	Linear distortions	333
8.6.3	Non-linear distortions	335
8.6.4	Recommendations on recording chain characteristics	335
8.7	Reproducibility assurance procedures	336
8.7.1	Description of recording conditions	336
8.7.2	References and standards	337
8.7.3	Recommendations on reproducibility	338
8.8	Analysis and conditioning tools for speech signals	339
8.8.1	Signal analysis and measurement	339
8.8.2	Signal conditioning	340
8.8.3	Recommendations on analysis and conditioning tools for speech signals	341

Part III: Spoken language system assessment 343

9	Assessment methodologies and experimental design	344
9.1	Introduction	344
9.1.1	How to read this chapter	344
9.1.2	Role of statistical analysis and experimentation in Language Engineering Standards (LES)	346

9.2	Statistical and experimental procedures for analysing data corpora	347
9.2.1	Statistical analysis	347
9.2.2	Populations, samples and other terminology	347
9.2.3	Sampling	348
9.2.4	Biases	348
9.2.5	Estimating sample means, proportions and variances	349
9.2.6	Hypothesis testing	354
9.3	Experimental procedures	360
9.3.1	Experimental selection of material	360
9.3.2	Segmentation	363
9.3.3	Classification	365
9.4	Assessing recognisers	366
9.4.1	Baseline performance	366
9.4.2	Progress	367
9.4.3	Functional adequacy and user acceptance	370
9.4.4	Methodology	372
9.4.5	Experimental design	373
9.5	Assessing speaker verification and recognition systems	373
9.5.1	Sampling rare events in speaker verification and recognition systems	374
9.5.2	Employing expert judgments to augment speaker verification and assessment for forensic aspects of speaker verification and recognition	374
9.6	Interactive dialogue systems	375
9.6.1	Wizard of Oz (WOZ)	375
9.6.2	Dialogue metrics	378
10	Assessment of recognition systems	381
10.1	Introduction	381
10.1.1	Classification of recognition systems	381
10.1.2	Speech quality and conditions	383
10.1.3	Capability profile versus requirement profile	384
10.1.4	Assessment purpose versus methodology	385
10.2	Definitions and nomenclature	385
10.2.1	The performance measure as percentage	385
10.2.2	Recognition score	386
10.2.3	Confusions	388
10.2.4	Vocabulary	388
10.2.5	Analysis of Variance design	389
10.3	Description of methodologies	389
10.3.1	Representative databases	389
10.3.2	Reference methods	390

10.3.3	Specific calibrated databases	391
10.3.4	Diagnostic methods with a specific vocabulary	391
10.3.5	Artificial test signals	392
10.4	Parameters	392
10.4.1	Pre-production parameters	392
10.4.2	Post-production parameters	393
10.4.3	Linguistic parameters	393
10.4.4	Recogniser specific parameters	393
10.4.5	Assessment parameters	394
10.5	Experimental design of small vocabulary word recognition . .	396
10.5.1	Technical set-up	396
10.5.2	Training	398
10.5.3	Test procedure	398
10.5.4	Scoring the results	401
10.5.5	Analysis of results	401
10.6	Experimental design of large vocabulary continuous speech recognition	401
10.6.1	Training material	402
10.6.2	Development test	405
10.6.3	Dry run	405
10.6.4	Test material selection	405
10.6.5	Evaluation protocol	407
10.6.6	Scoring method	407
11	Assessment of speaker verification systems	408
11.1	Presentation	408
11.1.1	Speaker classification tasks	408
11.1.2	General definitions	409
11.2	A taxonomy of speaker recognition systems	410
11.2.1	Task typology	410
11.2.2	Levels of text dependence	412
11.2.3	Interaction mode with the user	412
11.2.4	Definitions	413
11.2.5	Examples	415
11.3	Influencing factors	418
11.3.1	Speech quality	418
11.3.2	Temporal drift	419
11.3.3	Speech quantity and variety	419
11.3.4	Speaker population size and typology	420
11.3.5	Speaker purpose and other human factors	421
11.3.6	Recommendations	423
11.3.7	Example	426
11.4	Scoring procedures	428

11.4.1	Notation	428
11.4.2	Closed-set identification	431
11.4.3	Verification	436
11.4.4	Open-set identification	458
11.4.5	Recommendations	459
11.5	Comparative and indirect assessment	461
11.5.1	Reference systems	461
11.5.2	Human calibration	463
11.5.3	Transformation of speech databases	464
11.6	Applications, systems and products	464
11.6.1	Terminology	465
11.6.2	Typology of applications	466
11.6.3	Examples of speaker verification systems	469
11.6.4	Examples of speaker verification products	470
11.6.5	Alternative techniques	471
11.6.6	Conclusions	473
11.7	System and product assessment	473
11.7.1	System assessment	474
11.7.2	Product assessment	475
11.7.3	Recommendations	476
11.8	Forensic applications	477
11.8.1	Listener method	477
11.8.2	Spectrographic method	478
11.8.3	Semi-automatic method	479
11.8.4	Recommendations	479
11.9	Conclusions	480
12	Assessment of synthesis systems	481
12.1	Introduction	481
12.1.1	What are speech output systems?	481
12.1.2	Why speech output assessment?	482
12.1.3	Users of this chapter	483
12.2	Towards a taxonomy of assessment tasks and techniques	485
12.2.1	Glass box vs. black box	485
12.2.2	Laboratory vs. field	487
12.2.3	Linguistic vs. acoustic	488
12.2.4	Human vs. automated	488
12.2.5	Judgment vs. functional testing	489
12.2.6	Global vs. analytic assessment	490
12.3	Methodology	490
12.3.1	Subjects	491
12.3.2	Test procedures	494
12.3.3	Benchmarks	497

12.3.4	Reference conditions	497
12.3.5	Comparability across languages	501
12.4	Black box approach	502
12.4.1	Laboratory testing	502
12.4.2	Field testing	508
12.5	Glass box approach	511
12.5.1	Linguistic aspects	511
12.5.2	Acoustic aspects	518
12.6	Further developments in speech output testing	540
12.6.1	Introduction	540
12.6.2	Long-term strategy: Towards predictive tests	541
12.6.3	Linguistic testing: Creating test environments for lin- guistic interfaces	544
12.6.4	Acoustic testing: Developments for the near future	546
12.7	Conclusion: summary of test descriptions	550
12.7.1	SAM Standard Segmental Test	551
12.7.2	CLuster IDentification Test (CLID)	552
12.7.3	The Bellcore Test	553
12.7.4	Diagnostic Rhyme Test (DRT)	554
12.7.5	Modified Rhyme Test (MRT)	555
12.7.6	Haskins Syntactic Sentences	556
12.7.7	SAM Semantically Unpredictable Sentences (SUS)	557
12.7.8	Harvard Psychoacoustic Sentences	558
12.7.9	SAM Prosodic Form Test	559
12.7.10	SAM Prosodic Function Test	560
12.7.11	SAM Overall Quality Test	561
12.7.12	ITU-T Overall Quality Test	562
13	Assessment of interactive systems	564
13.1	Introduction	564
13.1.1	About this chapter	564
13.1.2	Reading guide	565
13.2	Interactive dialogue systems	567
13.2.1	Definitions	567
13.3	Specification and design	571
13.3.1	Design by intuition	573
13.3.2	Design by observation	578
13.3.3	Design by simulation	581
13.3.4	Iterative design methodology for spoken language di- alogue systems	591
13.3.5	Readings in interactive dialogue system specification	594
13.4	Evaluation	594
13.4.1	Background	595

13.4.2	Characterisation	597
13.4.3	Assessment framework	601
13.4.4	Recommendations on evaluation methodology	613
13.4.5	Readings in interactive dialogue system evaluation	614

Bibliographical references	617
----------------------------	-----

Part IV: Spoken language reference materials	653
--	-----

A	Character codes and computer readable alphabets	654
A.1	Introduction	654
A.2	ASCII	655
A.2.1	Problems	655
A.3	Multi-Byte encodings	656
A.3.1	ISO 10646	656
A.3.2	The Unicode standard	656
A.4	Translation of code tables	657
A.4.1	Specific vs. general converters	657
A.4.2	MIME	658
A.5	Computer readable phonetic alphabets	658
A.5.1	IPA subsets	659
A.5.2	Full IPA alphabet encoding	660
A.6	References	662
B	SAMPA computer readable phonetic alphabet	684
B.1	Introduction	684
B.1.1	Notation issues	684
B.1.2	Transcription	684
B.1.3	Coding	685
B.1.4	Further languages	687
B.2	SAMPA: Present status	688
B.3	The phonemic notation of individual languages	691
B.3.1	Bulgarian	692
B.3.2	Danish	694
B.3.3	Dutch	697
B.3.4	English	699
B.3.5	Estonian	703
B.3.6	French	705
B.3.7	German	708
B.3.8	Greek	711
B.3.9	Hungarian	713
B.3.10	Italian	715

B.3.11 Norwegian	718
B.3.12 Polish	720
B.3.13 Portuguese	721
B.3.14 Romanian	723
B.3.15 Spanish	724
B.3.16 Swedish	726
B.4 Levels of annotation and extension of SAMPA	728
B.4.1 SAMPA as a phonemic system	728
B.4.2 Detailed phonetic or acoustic annotation	728
B.5 SAMPROSA (SAM PROSodic Alphabet): symbol set	731
B.6 References	732
C SAM file formats	733
C.1 Introduction	733
C.2 Speech file and associated description file formats	733
C.3 Label file format	733
C.3.1 Label file header format	734
C.3.2 Label file body format	735
C.3.3 Label types	735
C.3.4 Label File example	736
C.4 Files used in a recording session using EUROPEC	738
C.4.1 Corpus File: CORPUS.DBF	738
C.4.2 Speaker File: SPEAKERS.DBF	740
C.4.3 Prompt file	741
C.5 Protocol description file	743
C.5.1 Codes used for presentation protocol (Protocol De- scription File)	743
C.5.2 Example of Protocol Description File	744
C.6 Recording conditions file	746
C.7 PRIVATE.PEC file	746
D SAM recording protocols	748
D.1 Definition of terms	748
D.2 Classification of general strategies for recording and prompting	749
D.2.1 Recording mode	749
D.3 Prompting style	749
D.3.1 Timing strategy	750
D.4 Recording protocol	750
D.4.1 Microphone	750
D.4.2 Other sensors	751
D.4.3 Speech data capture	751
D.4.4 Recording environment	752
D.4.5 Recording mode and prompting style	752

D.4.6	Recording control	753
D.4.7	Recording procedure	755
D.4.8	Integrity checks	756
D.4.9	Backup procedures	756
D.4.10	Retrieval procedures	756
D.4.11	Calibration	757
D.4.12	Inter site consistency and recording procedure verification	757
D.4.13	Collation of recordings	757
E	SAM software tools	759
E.1	DKISALA (Interactive Semi-Automatic Labelling Software)	759
E.2	ELSA (ESPRIT Labelling System Assessment software)	760
E.3	EUROPEC (European Program d'Enregistrement de Corpus)	761
E.4	PTS (Progiciel de Traitement de Signal)	761
E.5	RESAM	762
E.6	SAMITPRO (SAM Iterative Proportional Fitting)	763
E.7	SAM_REC0 Isolated Word Recogniser	764
E.8	SAM_SCOR (Sam Input Assessment Scoring Software)	765
E.9	SAM_SLM	766
E.10	SAM_SPEX (Speech Parameter Extractor)	767
E.11	SAMTRA (SAM TRanscription Analysis)	768
E.12	SOAP (Speech Output Assessment Package)	768
E.13	PTM (Parametric Test Manager)	769
F	EUROPEC recording tool	771
G	Digital storage media	774
G.1	Storage media dimensions	774
G.1.1	Permanence	774
G.1.2	Order of access	774
G.1.3	Other storage media characteristics	774
G.1.4	Compression	775
G.2	CD-ROM	775
G.2.1	Physical CD formats	776
G.2.2	Logical data organisation	776
G.2.3	CD-R software	777
G.2.4	CD-R hardware	777
G.2.5	Future developments	778
G.3	Summary	778
G.4	References	779
H	Database Management Systems (DBMSs)	780
H.1	Data model	781

H.1.1	Hierarchical data model	782
H.1.2	Network data model	782
H.1.3	Relational data model	782
H.1.4	Object-oriented data model	785
H.1.5	Deductive data model	785
H.2	Safe storage of data	786
H.3	Application-independent storage of data	786
H.3.1	Controlled access to data	787
H.4	Summary	787
H.5	References	788
I	Speech standards	789
I.1	Introduction	789
I.2	Signal theory recalls	789
I.3	Speech research	790
I.4	Computer hardware and software	794
I.5	Conclusion	795
I.6	Appendix: Useful anonymous ftp sites	796
J	EUROM-1 database overview	797
K	Polyphone project overview	799
K.1	Introduction	799
K.2	Reusable resources	800
K.3	The Dutch POLYPHONE corpus	801
K.3.1	Recording workstation	801
K.3.2	Speaker selection	801
K.3.3	The speech material	802
K.3.4	Postprocessing	803
K.4	Use of POLYPHONE in application development	803
K.4.1	Train time table information	803
K.4.2	Phone card services	806
K.4.3	Speaker verification	807
K.5	Requirements for future corpora	807
L	European speech resources	809
L.1	Overview	809
L.2	Criteria for assessment of the situation of Spoken Language Resources	810
L.2.1	Types and specificities of corpora	811
L.2.2	Actors in speech resource production	812
L.3	Summary of the current situation on a per language basis	813
L.3.1	DANISH (Denmark)	813
L.3.2	DUTCH (The Netherlands)	814

L.3.3	ENGLISH (United Kingdom)	814
L.3.4	FRENCH (France, Belgium, Switzerland)	815
L.3.5	GERMAN (Germany)	816
L.3.6	GREEK (Greece)	817
L.3.7	ITALIAN (Italy)	818
L.3.8	NORWEGIAN (Norway)	818
L.3.9	PORTUGUESE (Portugal)	819
L.3.10	SPANISH (Spain)	820
L.3.11	SWEDISH (Sweden)	820
L.4	General conclusions	821
L.5	Production costs	823
M	Transcription and documentation conventions for Speechdat	825
M.1	Introduction	825
M.1.1	Points of departure	825
M.2	Markings required for scoring	826
M.2.1	Case	826
M.2.2	Spelling	826
M.2.3	Number sequences	827
M.2.4	Letter sequences	827
M.2.5	Punctuation	828
M.2.6	Mispronunciations	828
M.2.7	Unintelligible words	828
M.2.8	Word fragments	828
M.2.9	Verbal deletions	828
M.2.10	Non-speech acoustic events	829
M.2.11	Prosodic annotation	830
M.2.12	Truncated waveforms	830
M.3	Assessment (optional)	831
M.4	Procedure	832
M.5	Speaker information	832
M.6	Some language-specific transcription options	832
M.6.1	French	832
M.6.2	English	833
M.6.3	Italian	834
N	The Bavarian Archive for Speech Signals	835
N.1	Introduction	835
N.2	General information	835
N.3	Corpora	835
N.4	Corpora in preparation	836
N.5	Other activities	837
N.6	Cooperations	837

N.7 Future plans	837
N.8 More information	837
Glossary	839
List of abbreviations	853
Index	859

List of Figures

1.1	The main structure of the EAGLES Group	2
1.2	The EAGLES workplan	3
1.3	Multimodal human-computer interface (HCI) including speech/language input/output	5
1.4	Working structure of the EAGLES Spoken Language Working Group	12
1.5	A model of the relationship between the applications of spo- ken language systems and the underlying technology	17
2.1	Spoken dialogue system	31
2.2	Capabilities versus requirements	33
2.3	Specifications versus capabilities	33
2.4	Technology specifications versus application requirements	34
2.5	Result presentation with "speaker selection" procedure	43
2.6	False acceptance versus false rejection	52
2.7	Error rate related to speech input duration	55
5.1	Levels of annotation	149
7.1	Bayes decision rule for speech recognition	240
7.2	Illustration of the decision problem for a three-word vocabulary	242
7.3	Perplexity as a function of b for absolute discounting with backing-off	261
7.4	Perplexity as a function of λ for linear discounting with backing-off	262
7.5	Bigram as a finite state grammar	268
7.6	Trigram as a finite state grammar	268
7.7	Parsing in the conventional form	269
7.8	Parsing in link grammar form	270
7.9	Example of a time alignment path	272
7.10	Bigram language model and search	273
7.11	Word recombination for a lexical tree	274
7.12	Simplified example of a word graph	276
7.13	Word pair approximation	278
7.14	Example of a word graph	280
8.1	Scheme of the communication chain	289
8.2	Typical polar patterns of various types of unidirectional mi- crophone	305
8.3	Closed room pressure zones	307
8.4	Typical comb filter structure	308

8.5	The minimum recording chain	311
8.6	Noise cancellation on balanced microphone lines	313
8.7	Technical principle of an electrical PDI	319
8.8	Augmented auditory subsystem of the SCATIS VE generator	328
9.1	Summary of decision structure for establishing what statisti- cal test to use for data	358
9.2	Activity associated with signal distribution	368
9.3	Activity associated with noise distribution	369
9.4	Activity level associated with signal and noise distribution . .	370
9.5	Schematic illustration of terms used in connection with speaker interruption patterns	380
10.1	Simplest set-up for interactive testing of a word recognition system	396
11.1	A typical ROC curve and its equal error rate	452
12.1	Relationships among dimensions involved in a taxonomy of speech output evaluation methods	486

List of Tables

2.1	Confusion matrix	42
2.2	Error presentation	49
2.3	False acceptance versus false rejection	51
2.4	Different types of error	53
2.5	Erlang B carried-traffic table	76
2.6	Erlang B carried-traffic table (continued)	77
2.7	Erlang B carried-traffic table (continued)	78
4.1	Possible layout for recording session information	133
5.1	Table of number-to-symbol correspondence for the <code>L^AT_EX</code> font <code>wsuipa11</code>	173
6.1	Examples of morphological coverage	190
6.2	Computer readable ASCII orthography for German	205
6.3	Frequently used symbols	226
6.4	VERBMOBIL diacritics	227
6.5	Extract from the VERBMOBIL pronunciation table	228
7.1	Effect of trigram model on recognition (Jelinek 1991)	244
7.2	Number of words and sentences in training and test data (vocabulary: about 20000 words)	258
7.3	Number of distinct and of singleton events for bigrams and trigrams	258
7.4	Perplexities for different language models	259
7.5	Algorithm for word clustering	265
7.6	One-pass algorithm for word graph construction (“single best” and “word graph”)	279
7.7	DP algorithm for search through word graph	280
10.1	Classification of speech recognition systems	381
10.2	Conditions of speech	383
11.1	Verification Phrase Construction for the TI Operational Voice Verification System (after Doddington)	416
11.2	Example of misclassification & mistrust rate scoring from a confusion matrix in speaker identification.	437
11.3	Example of confidence rank scoring in closed-set speaker identification	438
11.4	Genuine attempts	447
11.5	Impostor attempts against registered speakers	448

11.6	Impostor attempts from impostors	449
11.7	Biometric verifier	472
11.8	Important properties for designing a security system	472
11.9	Values for different biometric verifiers	472
12.1	Percentage of correct treatment of textual anomalies by CSTR text preprocessor (after Laver et al. 1988, pp. 12–15)	512
12.2	Evaluation results for text preprocessor TextScan (after Van Holsteijn 1993).	513
12.3	Percentage correct grapheme–phoneme conversion in seven EU languages in four types of materials (after Pols 1991, p. 394)	514
12.4	Some results obtained with the DRT by Pratt (1987)	521
12.5	Some results for the MRT (Logan et al. 1985)	522
12.6	Some results for the SAM Standard Segmental Test (Goldstein and Till 1992)	523
12.7	Some results for the Harvard Psychoacoustic Sentences and Haskins Syntactic Sentences	528
12.8	Percentage of correctly recognised emotions-by-rule in Dutch diphone synthesis (two diphone sets, obtained from different speakers) and in human speech (after Vroomen et al. 1993)	535
13.1	Summary of core comparative evaluation metrics	614
A.1	The KLATTBET phonetic alphabet	660
A.2	Machine Readable Phonetic Alphabet (MRPA)	661
A.3	IPA symbol chart	663
A.4	IPA number chart	664
A.5	IPA table (ordered by symbol)	665
A.6	IPA table (ordered by symbol) (continued)	666
A.7	IPA table (ordered by symbol)(continued)	667
A.8	IPA table (ordered by symbol) (continued)	668
A.9	IPA table (ordered by symbol) (continued)	669
A.10	IPA table (ordered by symbol) (continued)	670
A.11	IPA table (ordered by symbol) (continued)	671
A.12	IPA table (ordered by symbol) (continued)	672
A.13	IPA table (ordered by symbol) (continued)	673
A.14	IPA table (ordered by symbol) (continued)	674
A.15	IPA table (ordered by number)	675
A.16	IPA table (ordered by number) (continued)	676
A.17	IPA table (ordered by number) (continued)	677
A.18	IPA table (ordered by number) (continued)	678
A.19	IPA table (ordered by number) (continued)	679
A.20	IPA table (ordered by number) (continued)	680

A.21 IPA table (ordered by number) (continued)	681
A.22 IPA table (ordered by number) (continued)	682
A.23 IPA table (ordered by number) (continued)	683
B.1 Vowels	688
B.2 Consonants	689
B.3 Two character symbols	689
B.4 Boundary and prosodic features	690
B.5 Currently under discussion	690
B.6 Currently used in French work	690
G.1 Storage media classification	779
I.1 Sampling frequencies	790
M.1 CSELT mapping between letters and graphemes	834