

Contents

	<i>Preface</i>	<i>page</i> ix
	<i>Book features</i>	xii
	<i>Acknowledgements</i>	xv
1	Introduction	1
	1.1 Computers and audio	1
	1.2 Digital audio	3
	1.3 Capturing and converting sound	4
	1.4 Sampling	5
	1.5 Summary	6
	Bibliography	7
2	Basic audio processing	9
	2.1 Sound in MATLAB	10
	2.2 Normalisation	18
	2.3 Continuous audio processing	20
	2.4 Segmentation	24
	2.5 Analysis window sizing	32
	2.6 Visualisation	37
	2.7 Sound generation	44
	2.8 Summary	50
	Bibliography	50
	Questions	52
3	The human voice	54
	3.1 Speech production	55
	3.2 Characteristics of speech	57
	3.3 Types of speech	67
	3.4 Speech understanding	71
	3.5 Summary	82
	Bibliography	83
	Questions	83

vi	Contents	
4	The human auditory system	85
	4.1 Physical processes	85
	4.2 Perception	87
	4.3 Amplitude and frequency models	103
	4.4 Summary	107
	Bibliography	107
	Questions	108
5	Psychoacoustics	109
	5.1 Psychoacoustic processing	109
	5.2 Auditory scene analysis	112
	5.3 Psychoacoustic modelling	121
	5.4 Hermansky-style model	132
	5.5 MFCC model	134
	5.6 Masking effect of speech	137
	5.7 Summary	138
	Bibliography	138
	Questions	139
6	Speech communications	140
	6.1 Quantisation	140
	6.2 Parameterisation	148
	6.3 Pitch models	176
	6.4 Analysis-by-synthesis	182
	6.5 Perceptual weighting	191
	6.6 Summary	192
	Bibliography	192
	Questions	193
7	Audio analysis	195
	7.1 Analysis toolkit	196
	7.2 Speech analysis and classification	208
	7.3 Some examples of audio analysis	211
	7.4 Statistics and classification	213
	7.5 Analysing other signals	216
	7.6 Summary	220
	Bibliography	220
	Questions	221
8	Big data	223
	8.1 The rationale behind big data	225
	8.2 Obtaining big data	226

8.3	Classification and modelling	227
8.4	Summary of techniques	234
8.5	Big data applications	263
8.6	Summary	264
	Bibliography	264
	Questions	265
9	Speech recognition	267
9.1	What is speech recognition?	267
9.2	Voice activity detection and segmentation	275
9.3	Current speech recognition research	282
9.4	Hidden Markov models	288
9.5	ASR in practice	298
9.6	Speaker identification	302
9.7	Language identification	305
9.8	Diarization	308
9.9	Related topics	309
9.10	Summary	311
	Bibliography	311
	Questions	312
10	Advanced topics	314
10.1	Speech synthesis	314
10.2	Stereo encoding	324
10.3	Formant strengthening and steering	334
10.4	Voice and pitch changer	338
10.5	Statistical voice conversion	346
10.6	Whisper-to-speech conversion	347
10.7	Whisperisation	354
10.8	Super-audible speech	357
10.9	Summary	363
	Bibliography	364
	Questions	365
11	Conclusion	366
	<i>References</i>	370
	<i>Index</i>	379