Contents

About the Authors xix
Preface xxi
Preface to the Unfinished Manuscript of the Book xxiii
Introduction 1

1 How to Study and Develop Communication Acoustics 7
   1.1 Domains of Knowledge 7
   1.2 Methodology of Research and Development 8
   1.3 Systems Approach to Modelling 10
   1.4 About the Rest of this Book 12
   1.5 Focus of the Book 12
   1.6 Intended Audience 13
   References 14

2 Physics of Sound 15
   2.1 Vibration and Wave Behaviour of Sound 15
      2.1.1 From Vibration to Waves 16
      2.1.2 A Simple Vibrating System 16
      2.1.3 Resonance 18
      2.1.4 Complex Mass–Spring Systems 19
      2.1.5 Modal Behaviour 20
      2.1.6 Waves 21
   2.2 Acoustic Measures and Quantities 23
      2.2.1 Sound and Voice as Signals 23
      2.2.2 Sound Pressure 24
      2.2.3 Sound Pressure Level 24
      2.2.4 Sound Power 25
      2.2.5 Sound Intensity 25
      2.2.6 Computation with Amplitude and Level Quantities 25
## Contents

### 2.3 Wave Phenomena
- 2.3.1 Spherical Waves
- 2.3.2 Plane Waves and the Wave Field in a Tube
- 2.3.3 Wave Propagation in Solid Materials
- 2.3.4 Reflection, Absorption, and Refraction
- 2.3.5 Scattering and Diffraction
- 2.3.6 Doppler Effect

### 2.4 Sound in Closed Spaces: Acoustics of Rooms and Halls
- 2.4.1 Sound Field in a Room
- 2.4.2 Reverberation
- 2.4.3 Sound Pressure Level in a Room
- 2.4.4 Modal Behaviour of Sound in a Room
- 2.4.5 Computational Modelling of Closed Space Acoustics

### 3 Signal Processing and Signals
- 3.1 Signals
  - 3.1.1 Sounds as Signals
  - 3.1.2 Typical Signals
- 3.2 Fundamental Concepts of Signal Processing
  - 3.2.1 Linear and Time-Invariant Systems
  - 3.2.2 Convolution
  - 3.2.3 Signal Transforms
  - 3.2.4 Fourier Analysis and Synthesis
  - 3.2.5 Spectrum Analysis
  - 3.2.6 Time–Frequency Representations
  - 3.2.7 Filter Banks
  - 3.2.8 Auto- and Cross-Correlation
  - 3.2.9 Cepstrum
- 3.3 Digital Signal Processing (DSP)
  - 3.3.1 Sampling and Signal Conversion
  - 3.3.2 Z Transform
  - 3.3.3 Filters as LTI Systems
  - 3.3.4 Digital Filtering
  - 3.3.5 Linear Prediction
  - 3.3.6 Adaptive Filtering
- 3.4 Hidden Markov Models
- 3.5 Concepts of Intelligent and Learning Systems

### 4 Electroacoustics and Responses of Audio Systems
- 4.1 Electroacoustics
  - 4.1.1 Loudspeakers
  - 4.1.2 Microphones
Contents

4.2 Audio System Responses
   4.2.1 Measurement of System Response 71
   4.2.2 Ideal Reproduction of Sound 72
   4.2.3 Impulse Response and Magnitude Response 72
   4.2.4 Phase Response 74
   4.2.5 Non-Linear Distortion 75
   4.2.6 Signal-to-Noise Ratio 76

4.3 Response Equalization 76
Summary 77
Further Reading 78
References 78

5 Human Voice

5.1 Speech Production 79
   5.1.1 Speech Production Mechanism 80
   5.1.2 Vocal Folds and Phonation 80
   5.1.3 Vocal and Nasal Tract and Articulation 82
   5.1.4 Lip Radiation Measurements 84

5.2 Units and Notation of Speech used in Phonetics 84
   5.2.1 Vowels 86
   5.2.2 Consonants 86
   5.2.3 Prosody and Suprasegmental Features 88

5.3 Modelling of Speech Production 90
   5.3.1 Glottal Modelling 92
   5.3.2 Vocal Tract Modelling 92
   5.3.3 Articulatory Synthesis 94
   5.3.4 Formant Synthesis 95

5.4 Singing Voice 96
Summary 96
Further Reading 97
References 97

6 Musical Instruments and Sound Synthesis

6.1 Acoustic Instruments 99
   6.1.1 Types of Musical Instruments 99
   6.1.2 Resonators in Instruments 100
   6.1.3 Sources of Excitation 102
   6.1.4 Controlling the Frequency of Vibration 103
   6.1.5 Combining the Excitation and Resonant Structures 104

6.2 Sound Synthesis in Music 104
   6.2.1 Envelope of Sounds 105
   6.2.2 Synthesis Methods 106
   6.2.3 Synthesis of Plucked String Instruments with a One-Dimensional Physical Model 107

Summary 108
Further Reading 108
References 108
7 Physiology and Anatomy of Hearing

7.1 Global Structure of the Ear

7.2 External Ear

7.3 Middle Ear

7.4 Inner Ear

7.4.1 Structure of the Cochlea

7.4.2 Passive Cochlear Processing

7.4.3 Active Function of the Cochlea

7.4.4 The Inner Hair Cells

7.4.5 Cochlear Non-Linealities

7.5 Otoacoustic Emissions

7.6 Auditory Nerve

7.6.1 Information Transmission using the Firing Rate

7.6.2 Phase Locking

7.7 Auditory Nervous System

7.7.1 Structure of the Auditory Pathway

7.7.2 Studying Brain Function

7.8 Motivation for Building Computational Models of Hearing

Summary

Further Reading

References

8 The Approach and Methodology of Psychoacoustics

8.1 Sound Events versus Auditory Events

8.2 Psychophysical Functions

8.3 Generation of Sound Events

8.3.1 Synthesis of Sound Signals

8.3.2 Listening Set-up and Conditions

8.3.3 Steering Attention to Certain Details of An Auditory Event

8.4 Selection of Subjects for Listening Tests

8.5 What are We Measuring?

8.5.1 Thresholds

8.5.2 Scales and Categorization of Percepts

8.5.3 Numbering Scales in Listening Tests

8.6 Tasks for Subjects

8.7 Basic Psychoacoustic Test Methods

8.7.1 Method of Constant Stimuli

8.7.2 Method of Limits

8.7.3 Method of Adjustment

8.7.4 Method of Tracking

8.7.5 Direct Scaling Methods

8.7.6 Adaptive Staircase Methods

8.8 Descriptive Sensory Analysis

8.8.1 Verbal Elicitation

8.8.2 Non-Verbal Elicitation

8.8.3 Indirect Elicitation
## Contents

8.9 Psychoacoustic Tests from the Point of View of Statistics 149  
Summary 149  
Further Reading 150  
References 150

9 Basic Function of Hearing 153

9.1 Effective Hearing Area 153  
9.1.1 Equal Loudness Curves 155  
9.1.2 Sound Level and its Measurement 156

9.2 Spectral Masking 156  
9.2.1 Masking by Noise 157  
9.2.2 Masking by Pure Tones 159  
9.2.3 Masking by Complex Tones 159  
9.2.4 Other Masking Phenomena 161

9.3 Temporal Masking 161

9.4 Frequency Selectivity of Hearing 163  
9.4.1 Psychoacoustic Tuning Curves 164  
9.4.2 ERB Bandwidths 166  
9.4.3 Bark, ERB, and Greenwood Scales 167

Summary 169  
Further Reading 169  
References 169

10 Basic Psychoacoustic Quantities 171

10.1 Pitch 171  
10.1.1 Pitch Strength and Frequency Range 171

10.1.2 JND of Pitch 172

10.1.3 Pitch Perception versus Duration of Sound 173

10.1.4 Mel Scale 174

10.1.5 Logarithmic Pitch Scale and Musical Scale 175

10.1.6 Detection Threshold of Pitch Change and Frequency Modulation 176

10.1.7 Pitch of Coloured Noise 176

10.1.8 Repetition Pitch 177

10.1.9 Virtual Pitch 178

10.1.10 Pitch of Non-Harmonic Complex Sounds 178

10.1.11 Pitch Theories 178

10.1.12 Absolute Pitch 179

10.2 Loudness 179  
10.2.1 Loudness Determination Experiments 179

10.2.2 Loudness Level 180

10.2.3 Loudness of a Pure Tone 180

10.2.4 Loudness of Broadband Signals 182

10.2.5 Excitation Pattern, Specific Loudness, and Loudness 183

10.2.6 Difference Threshold of Loudness 185

10.2.7 Loudness versus Duration of Sound 187
10.3 Timbre

10.3.1 Timbre of Steady-State Sounds

10.3.2 Timbre of Sound Including Modulations

10.4 Subjective Duration of Sound

Summary

Further Reading

References

11 Further Analysis in Hearing

11.1 Sharpness

11.2 Detection of Modulation and Sound Onset

11.2.1 Fluctuation Strength

11.2.2 Impulsiveness

11.3 Roughness

11.4 Tonality

11.5 Discrimination of Changes in Signal Magnitude and Phase Spectra

11.5.1 Adaptation to the Magnitude Spectrum

11.5.2 Perception of Phase and Time Differences

11.6 Psychoacoustic Concepts and Music

11.6.1 Sensory Consonance and Dissonance

11.6.2 Intervals, Scales, and Tuning in Music

11.6.3 Rhythm, Tempo, Bar, and Measure

11.7 Perceptual Organization of Sound

11.7.1 Segregation of Sound Sources

11.7.2 Sound Streaming and Auditory Scene Analysis

Summary

Further Reading

References

12 Spatial Hearing

12.1 Concepts and Definitions for Spatial Hearing

12.1.1 Basic Concepts

12.1.2 Coordinate Systems for Spatial Hearing

12.2 Head-Related Acoustics

12.3 Localization Cues

12.3.1 Interaural Time Difference

12.3.2 Interaural Level Difference

12.3.3 Interaural Coherence

12.3.4 Cues to Resolve the Direction on the Cone of Confusion

12.3.5 Interaction Between Spatial Hearing and Vision

12.4 Localization Accuracy

12.4.1 Localization in the Horizontal Plane

12.4.2 Localization in the Median Plane

12.4.3 3D Localization

12.4.4 Perception of the Distribution of a Spatially Extended Source
12.5 Directional Hearing in Enclosed Spaces
   12.5.1 Precedence Effect
   12.5.2 Adaptation to the Room Effect in Localization
12.6 Binaural Advantages in Timbre Perception
   12.6.1 Binaural Detection and Unmasking
   12.6.2 Binaural Decolouration
12.7 Perception of Source Distance
   12.7.1 Cues for Distance Perception
   12.7.2 Accuracy of Distance Perception
Summary
Further Reading
References

13 Auditory Modelling
   13.1 Simple Psychoacoustic Modelling with DFT
      13.1.1 Computation of the Auditory Spectrum through DFT
   13.2 Filter Bank Models
      13.2.1 Modelling the Outer and Middle Ear
      13.2.2 Gammatone Filter Bank and Auditory Nerve Responses
      13.2.3 Level-Dependent Filter Banks
      13.2.4 Envelope Detection and Temporal Dynamics
   13.3 Cochlear Models
      13.3.1 Basilar Membrane Models
      13.3.2 Hair-Cell Models
   13.4 Modelling of Higher-Level Systemic Properties
      13.4.1 Analysis of Pitch and Periodicity
      13.4.2 Modelling of Loudness Perception
   13.5 Models of Spatial Hearing
      13.5.1 Delay-Network-Based Models of Binaural Hearing
      13.5.2 Equalization Cancellation and ILD Models
      13.5.3 Count-Comparison Models
      13.5.4 Models of Localization in the Median Plane
   13.6 Matlab Examples
      13.6.1 Filter-Bank Model with Autocorrelation-Based Pitch Analysis
      13.6.2 Binaural Filter-Bank Model with Cross-Correlation-Based ITD Analysis
Summary
Further Reading
References

14 Sound Reproduction
   14.1 Need for Sound Reproduction
   14.2 Audio Content Production
   14.3 Listening Set-ups
      14.3.1 Loudspeaker Set-ups
      14.3.2 Listening Room Acoustics
## Contents

### 14.4 Recording Techniques
- 14.4.1 Monophonic Techniques
- 14.4.2 Spot Microphone Technique
- 14.4.3 Coincident Microphone Techniques for Two-Channel Stereophony
- 14.4.4 Spaced Microphone Techniques for Two-Channel Stereophony
- 14.4.5 Spaced Microphone Techniques for Multi-Channel Loudspeaker Systems
- 14.4.6 Coincident Recording for Multi-Channel Set-up with Ambisonics
- 14.4.7 Non-Linear Time–Frequency-domain Reproduction of Spatial Sound

### 14.5 Virtual Source Positioning
- 14.5.1 Amplitude Panning
- 14.5.2 Amplitude Panning in a Stereophonic Set-up
- 14.5.3 Amplitude Panning in Horizontal Multi-Channel Loudspeaker Set-ups
- 14.5.4 3D Amplitude Panning
- 14.5.5 Virtual Source Positioning using Ambisonics
- 14.5.6 Wave Field Synthesis
- 14.5.7 Time Delay Panning
- 14.5.8 Synthesizing the Width of Virtual Sources

### 14.6 Binaural Techniques
- 14.6.1 Listening to Binaural Recordings with Headphones
- 14.6.2 HRTF Processing for Headphone Listening
- 14.6.3 Virtual Listening of Loudspeakers with Headphones
- 14.6.4 Headphone Listening to Two-Channel Stereophonic Content
- 14.6.5 Binaural Techniques with Cross-Talk-Cancelled Loudspeakers

### 14.7 Digital Audio Effects
- 14.8 Reverberators
  - 14.8.1 Using Room Impulse Responses in Reverberators
  - 14.8.2 DSP Structures for Reverberators

### Summary
- 15 Time–Frequency-domain Processing and Coding of Audio
  - 15.1 Basic Techniques and Concepts for Time–Frequency Processing
    - 15.1.1 Frame-Based Processing
    - 15.1.2 Downsampled Filter-Bank Processing
    - 15.1.3 Modulation with Tone Sequences
    - 15.1.4 Aliasing
  - 15.2 Time–Frequency Transforms
    - 15.2.1 Short-Time Fourier Transform (STFT)
    - 15.2.2 Alias-Free STFT
    - 15.2.3 Modified Discrete Cosine Transform (MDCT)
## Contents

15.2.4 Pseudo-Quadrature Mirror Filter (PQMF) Bank 323  
15.2.5 Complex QMF 323  
15.2.6 Sub-Sub-Band Filtering of the Complex QMF Bands 325  
15.2.7 Stochastic Measures of Time–Frequency Signals 325  
15.2.8 Decorrelation 327  
15.3 Time–Frequency-Domain Audio-Processing Techniques 328  
15.3.1 Masking-Based Audio Coding 328  
15.3.2 Audio Coding with Spectral Band Replication 328  
15.3.3 Parametric Stereo, MPEG Surround, and Spatial Audio Object Coding 329  
15.3.4 Stereo Upmixing and Enhancement for Loudspeakers and Headphones 330  

Summary 332  
Further Reading 332  
References 332

16 Speech Technologies 335  
16.1 Speech Coding 336  
16.2 Text-to-Speech Synthesis 338  
16.2.1 Early Knowledge-Based Text-to-Speech (TTS) Synthesis 339  
16.2.2 Unit-Selection Synthesis 340  
16.2.3 Statistical Parametric Synthesis 342  
16.3 Speech Recognition 345  

Summary 346  
Further Reading 347  
References 347

17 Sound Quality 349  
17.1 Historical Background of Sound Quality 350  
17.2 The Many Facets of Sound Quality 351  
17.3 Systemic Framework for Sound Quality 352  
17.4 Subjective Sound Quality Measurement 353  
17.4.1 Mean Opinion Score 353  
17.4.2 MUSHRA 354  
17.5 Audio Quality 356  
17.5.1 Monaural Quality 356  
17.5.2 Perceptual Measures and Models for Monaural Audio Quality 356  
17.5.3 Spatial Audio Quality 359  
17.6 Quality of Speech Communication 360  
17.6.1 Subjective Methods and Measures 361  
17.6.2 Objective Methods and Measures 362  
17.7 Measuring Speech Understandability with the Modulation Transfer Function 363  
17.7.1 Modulation Transfer Function 363  
17.7.2 Speech Transmission Index STI 367
# Contents

17.7.3  STI and Speech Intelligibility 368  
17.7.4  Practical Measurement of STI 369  
17.8  Objective Speech Quality Measurement for Telecommunication 370  
17.8.1  General Speech Quality Measurement Techniques 371  
17.8.2  Measurement of the Perceptual Effect of Background Noise 372  
17.8.3  Measurement of the Perceptual Effect of Echoes 373  
17.9  Sound Quality in Auditoria and Concert Halls 374  
17.9.1  Subjective Measures 374  
17.9.2  Objective Measures 375  
17.9.3  Percentage of Consonant Loss 377  
17.10  Noise Quality 377  
17.11  Product Sound Quality 378  
Summary 380  
Further Reading 380  
References 380  

18  Other Audio Applications 383  
18.1  Virtual Reality and Game Audio Engines 383  
18.2  Sonic Interaction Design 386  
18.3  Computational Auditory Scene Analysis, CASA 387  
18.4  Music Information Retrieval 387  
18.5  Miscellaneous Applications 389  
Summary 390  
Further Reading 390  
References 390  

19  Technical Audiology 393  
19.1  Hearing Impairments and Disabilities 393  
19.1.1  Key Terminology 394  
19.1.2  Classification of Hearing Impairments 395  
19.1.3  Causes for Hearing Impairments 396  
19.2  Symptoms and Consequences of Hearing Impairments 396  
19.2.1  Hearing Threshold Shift 397  
19.2.2  Distortion and Decrease in Discrimination 398  
19.2.3  Speech Communication Problems 400  
19.2.4  Tinnitus 400  
19.3  The Effect of Noise on Hearing 401  
19.3.1  Noise 401  
19.3.2  Formation of Noise-Induced Hearing Loss 402  
19.3.3  Temporary Threshold Shift 402  
19.3.4  Hearing Protection 404  
19.4  Audiometry 405  
19.4.1  Pure-Tone Audiometry 405  
19.4.2  Bone-Conduction Audiometry 406  
19.4.3  Speech Audiometry 406  
19.4.4  Sound-Field Audiometry 407