

Communication Acoustics An Introduction to Speech, Audio and Psychoacoustics

Abramo Abessi



Communication Acoustics and Introduction to Speech, Audio and Psychoacoustics

Editor

Abramo Adessi



scitus
academics

Communication Acoustics An Introduction to Speech, Audio and Psychoacoustics

Edited by **Abramo Adessi**

ISBN: 978-1-68117-112-8

Library of Congress Control Number: 2015954091

© 2016 by

SCITUS Academics LLC,

www.scitusacademics.com

Box No. 4766, 616 Corporate Way,

Suite 2, Valley Cottage,

NY 10989

This book contains information obtained from highly regarded resources. Copyright for individual articles remains with the authors as indicated. All chapters are distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Notice

Reasonable efforts have been made to publish reliable data and views articulated in the chapters are those of the individual contributors, and not necessarily those of the editors or publishers. Editors or publishers are not responsible for the accuracy of the information in the published chapters or consequences of their use. The publisher believes no responsibility for any damage or grievance to the persons or property arising out of the use of any materials, instructions, methods or thoughts in the book. The editors and the publisher have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission has not been obtained. If any copyright holder has not been acknowledged, please write to us so we may rectify.

Printed in United States of America on Acid Free Paper

Communication Acoustics an Introduction to Speech, Audio and Psychoacoustics

Preface

The communication channel, in communication acoustics, comprises of a sound source, a channel (acoustic and/or electric) and finally the receiver: the human auditory system, a complex and intricate system that forms the way sound is heard. Consequently, when developing techniques in communication acoustics, such as in speech, audio and aided hearing, it is significant to understand the time–frequency–space resolution of hearing.

This book gives an introduction to the fields which concern some kind of communication channel having the human as listener in the end; the fields together are named as “communication acoustics”.

This book conveys to engineering students and researchers alike the relevant knowledge about the nature of acoustics, sound and hearing that will empower them to develop new technologies in this area from end to end getting a thorough understanding of how sound and hearing works. It covers the multidisciplinary area of acoustics, hearing, psychoacoustics, signal processing, speech and sound quality and is appropriate for senior undergraduate and graduate courses related to audio communication systems. It discusses the technologies and applications for sound synthesis and reproduction, and for speech and audio quality evaluation.

Table of Contents

Chapter 1:	On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common.....	1
ABSTRACT		1
INTRODUCTION		2
MATERIALS AND METHODS		3
Emotion Model		3
Databases.....		4
Spontaneous Emotion in Speech: the VAM Corpus.....		6
Emotion in Music: Now that’s what i Call Music (NTWICM) Database.....		7
Emotion in Sound events: Emotional Sound Database.....		7
Extraction of Acoustic Descriptors.....		10
RESULTS.....		12
Feature Relevance		12
Automatic Classification Experiments		19
DISCUSSION		23
ACKNOWLEDGMENTS		25
REFERENCES		26
CITATION		29
 Chapter 2:	 Acoustics and Vibro-Acoustics Applied in Space Industry.....	 31
INTRODUCTION		31
EXTERNAL ON BOARD MICROPHONES INSTALLATION DEVICES		33
Mathematical Models.....		36
Methodology for the External Measured SPL Correction.....		40

Table of Contents

VIBRO-ACOUSTIC MODELLING OF PAYLOAD FAIRINGS (PLF)	42
Model Description	43
Modelling Methodology	44
Fem Structural Meshes	49
Fem and Bem Acoustic Meshes	50
Model Excitation	51
Sea Fairing Vibro-Acoustic Model	51
Analysis Results	53
Model Validation	55
COMBUSTION INSTABILITIES OF LIQUID PROPELLED ROCKET ENGINES DUE TO CHAMBER COMBUSTION ACOUSTICS	62
Experimental Acoustic Modal Analysis (Eama)	63
Helmholtz Resonator	64
Finite Element Model	67
Experimental Setup	67
Hr Design	69
Results	70
CONCLUSIONS	72
REFERENCES	74
CITATION	76

Chapter 3: Surface Acoustic Wave Based Magnetic Sensors..... 77

INTRODUCTION	77
SAW BASED PASSIVE SENSORS	79
Basics of Saw Devices	80
Basic Design Concepts of Passive Saw Sensors	81
PASSIVE AND REMOTE SAW-BASED MAGNETIC SENSORS	83
Magnetostrictivesaw Devices	83
Saw-Gmi Sensors	86
POTENTIAL APPLICATIONS	102
CONCLUSION	104
REFERENCES	105
CITATION	109

Chapter 4: Auditory Perceptual Objects as Generative Models: Setting the Stage for Communication by Sound..... 111

ABSTRACT	111
COMMUNICATION CHANNELS	112

THE BUILDING BRICKS: REGULARITY, DEVIANCE, PREDICTIVE INFORMATION PROCESSING	115
AN OVERVIEW OF DETECTING NEW INFORMATION IN THE AUDITORY MODALITY	118
Forming Auditory Sensory Memory Representations	120
Predictive Model	121
Comparing Model Predictions with the Sensory Representation of the Current Sound	122
Evaluation	123
INITIAL BUILD-UP OF AN AUDITORY MODEL	124
Simple Regularities	124
AUDITORY REGULARITY REPRESENTATIONS, AUDITORY STREAMS, AUDITORY PERCEPTUAL OBJECTS	135
HOW AERS WORKS WHEN THE MODEL HAS BEEN SET UP	140
Simultaneous Stream Segregation	140
Temporal Heuristic Analysis	142
Competition and Establishing the Perceived Sound Organization	143
Finalizing Feature Integration	145
Deviance Detection	147
Maintenance of the Model	148
COMPARISON WITH EXISTING MODELS OF PREDICTIVE PROCESSING IN PERCEPTION	151
AERS ISSUES RELATED TO COMMUNICATION BY SOUND	153
LIMITATIONS OF AERS AND FUTURE DIRECTIONS	155
ACKNOWLEDGMENTS	157
REFERENCES	158
CITATION	177

Chapter 5: **Ray Trace Modeling of Underwater Sound Propagation..... 179**

INTRODUCTION	179
THEORY OF RAY ACOUSTICS	180
A RECIPE FOR TRACING OF RAYS	186
EIGENRAY DETERMINATION	188
ACOUSTIC ABSORPTION IN SEA WATER	190
BOUNDARY CONDITIONS AT THE SURFACE AND BOTTOM INTERFACES	191
SYNTHESIZING THE FREQUENCY DOMAIN TRANSFER FUNCTION AND THE TIME RESPONSES	194
SPECIAL CONSIDERATIONS	198
Frequency of Applications	198

Table of Contents

Caustics and Turning Points	199
The Principle of Reciprocity and Its Validity in Ray Modeling.....	200
The Validity of Using Plane Wave Reflection Coefficients	201
Bench Marking Ray Modeling	203
CASE STUDIES	204
Seasonal Variations of Communication Links	204
Seismic Noise Propagation.....	208
SUMMARY	211
REFERENCES	212
CITATION	213
 Chapter 6: Acoustics in Optical Fiber.....	215
INTRODUCTION	215
Acousto-Optic Tunable Filter	216
Tuning of Attenuation Depth.....	220
Tuning of 3-Db Bandwidth	223
Double-Pass Configuration	226
GAIN FLATTENING FILTER.....	227
CONCLUSION	233
REFERENCES	234
CITATION	236
 Chapter 7: Cascading Multi-Hop Reservation and Transmission in Underwater Acoustic Sensor Networks.....	237
ABSTRACT	237
INTRODUCTION	238
PROBLEM STATEMENTS	242
Space-Time Uncertainty.....	242
Hidden-Node Problem in UWSNs	243
PROPOSED CMRT PROTOCOL.....	244
System Description	244
Definition of States	245
Length of the Silence State	246
Channel Occupancy Priority.....	247
Data Transmission Using the Packet-Train Method	250
Handshaking Triggering and Back-Off Algorithm.....	252
SIMULATIONS AND RESULTS.....	253
Simulation Model.....	253
Simulation Results	254
Comparison of CMRT with Other MAC Protocols.....	256

Analysis of Hop-Delay	258
Effects of Inter-Nodal Distance	259
CONCLUSIONS	261
AUTHOR CONTRIBUTIONS.....	261
REFERENCES	262
CITATION	263
Index.....	265

Chapter 1

On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common

*Felix Weninger¹, Florian Eyben¹, Björn W. Schuller^{1,2}, Marcello Mortillaro²
and Klaus R. Scherer²*

¹Machine Intelligence and Signal Processing Group, Mensch-Maschine-Kommunikation, Technische Universität München, Munich, Germany

²Centre Interfacultaire en Sciences Affectives, Université de Genève, Geneva, Switzerland

ABSTRACT

Without doubt, there is emotional information in almost any kind of sound received by humans every day: be it the affective state of a person transmitted by means of speech; the emotion intended by a composer while writing a musical piece, or conveyed by a musician while performing it; or the affective state connected to an acoustic event occurring in the environment, in the soundtrack of a movie, or in a radio play. In the field of affective computing, there is currently some loosely connected research concerning either of these phenomena, but a holistic computational model of affect in sound is still lacking. In turn, for tomorrow's pervasive technical systems, including affective companions and robots, it is expected to be highly beneficial to understand the affective dimensions of "the sound that something makes," in order to evaluate the system's auditory environment and its own audio output. This article aims at a first step toward a holistic computational model: starting from standard acoustic feature extraction schemes in the domains of speech, music, and sound analysis, we interpret the worth of

individual features across these three domains, considering four audio databases with observer annotations in the arousal and valence dimensions. In the results, we find that by selection of appropriate descriptors, cross-domain arousal, and valence regression is feasible achieving significant correlations with the observer annotations of up to 0.78 for arousal (training on sound and testing on enacted speech) and 0.60 for valence (training on enacted speech and testing on music). The high degree of cross-domain consistency in encoding the two main dimensions of affect may be attributable to the co-evolution of speech and music from multimodal affect bursts, including the integration of nature sounds for expressive effects.

INTRODUCTION

Without doubt, emotional expressivity in sound is one of the most important methods of human communication. Not only human speech, but also music and ambient sound events carry emotional information. This information is transmitted by modulation of the acoustics and decoded by the receiver – a human conversation partner, the audience of a concert, or a robot or automated dialog system. By that, the concept of emotion that we consider in this article is the one of consciously conveyed emotion (in contrast, for example, to the “true” emotion of a human related to biosignals such as heart rate). In speech, for example, a certain affective state can be transmitted through a change in vocal parameters, e.g., by adjusting fundamental frequency and loudness (Scherer et al., 2003). In music, we consider the emotion intended by the composer of a piece – and by that, the performing artist(s) as actor(s) realizing an emotional concept such as “happiness” or “sadness.” This can manifest through acoustic parameters such as tempo, dynamics (*forte/piano*), and instrumentation (Schuller et al., 2010). In contrast to earlier research on affect recognition from singing (e.g., Daido et al., 2011), we focus on polyphonic music – by that adding the instrumentation as a major contribution to expressivity. As a connection between music and speech emotion, for example, the effect of musical training on human emotion recognition has been highlighted in related work (Nilsson and Sundberg, 1985; Thompson et al., 2004). Lastly, also the concept of affect in sound adopted in this article is motivated by the usage of (ambient) sounds as a method of communication – to elicit an

intended emotional response in the audience of a movie, radio play, or in the users of a technical system with auditory output.

In the field of affective computing, there is currently some loosely connected research concerning either of these phenomena (Schuller et al., 2011a; Drossos et al., 2012; Yang and Chen, 2012). Despite a number of perception studies suggesting overlap in the relevant acoustic parameters (e.g., Ilie and Thompson, 2006), a holistic computational model of affect in general sound is still lacking. In turn, for tomorrow's technical systems, including affective companions and robots, it is expected to be highly beneficial to understand the affective dimensions of "the sound that something makes," in order to evaluate the system's auditory environment and its own audio output.

In order to move toward such a unified framework for affect analysis, we consider feature relevance analysis and automatic regression with respect to continuous observer ratings of the main dimensions of affect, arousal, and valence, across speech, music, and ambient sound events. Thereby, on the feature side, we restrict ourselves to non-symbolic acoustic descriptors, thus eliminating more domain-specific higher-level concepts such as linguistics, chords, or key. In particular, we use a well proven set of "low-level" acoustic descriptors for paralinguistic analysis of speech (cf. Section 2.3). Then, we address the importance of acoustic descriptors for the automatic recognition of continuous arousal and valence in a "cross-domain" setting. We show that there exist large commonalities but also strong differences in the worth of individual descriptors for emotion prediction in the various domains. Finally, we carry out experiments with automatic regression on a selected set of "generic acoustic emotion descriptors."

MATERIALS AND METHODS

Emotion Model

Let us first clarify the model of emotion employed in this article. There is a debate in the field on which type of model to adopt for emotion differentiation: discrete (categorical) or dimensional (e.g., Mortillaro et al., 2012). We believe that these approaches are highly complementary. It has been copiously shown that discrete emotions in higher dimensional space can be mapped parsimoniously into lower dimensional space. Most

frequently, the two dimensions valence and arousal are chosen, although it can be shown that affective space is best structured by four dimensions – adding power and novelty to valence and arousal (Fontaine et al., 2007). Whether to choose a categorical or dimensional approach is thus dependent on the respective research context and the specific goals. Here, we chose a valence \times arousal dimensional approach because of the range of affective phenomena underlying our stimuli. In addition for some of our stimulus sets only dimensional annotations were available.

Databases

Let us now start the technical discussion in this article by a brief introduction of the data sets used in the present study on arousal and valence in speech, music, and sound. The collection of emotional audio data for the purpose of automatic analysis has often been driven by computer engineering. This is particularly true for speech data – considering applications, for example, in human-computer interaction. This has led to large databases of spontaneous emotion expression, for example, emotion in child-robot interaction (Steidl, 2009) or communication with virtual humans (McKeown et al., 2012), which are however limited to specific domains. In contrast, there are data sets from controlled experiments, featuring, for example, emotions expressed (“enacted”) by professional actors, with restricted linguistic content (e.g., phonetically balanced pseudo sentences) with the goal to allow for domain-independent analysis of the variation of vocal parameters (Burkhardt et al., 2005; Bänziger et al., 2012). In the case of polyphonic music, data sets are mostly collected with (commercial) software applications in mind – for example, categorization of music databases on end-user devices (“music mood recognition”; Yang and Chen, 2012). Finally, emotion analysis of general sounds has been attempted only recently (Sundaram and Schleicher, 2010; Drossos et al., 2012; Schuller et al., 2012). In this light, we selected the following databases for our analysis: the Geneva Multimodal Emotion Portrayals (GEMEP) set as an example for enacted emotional speech; the Vera am Mittag (VAM) database as an example for spontaneous emotional speech “sampled” from a “real-life” context; the “Now That’s What I Call Music” (NTWICM) database for mood recognition in popular music; and the recently introduced emotional sounds database.