

SPRINGER BRIEFS IN ELECTRICAL AND
COMPUTER ENGINEERING · SPEECH TECHNOLOGY

Koteswara Rao Anne
Swarna Kuchibhotla
Hima Deepthi Vankayalapati

Acoustic Modeling for Emotion Recognition

SpringerBriefs in Electrical and Computer Engineering

SpringerBriefs in Speech Technology

Studies in Speech Signal Processing, Natural Language Understanding, and Machine Learning

Series Editor

Amy Neustein

Fort Lee, New Jersey, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include: timely report of state-of-the art analytical techniques, a bridge between new research results, as published in journal articles, and a contextual literature review, a snapshot of a hot or emerging topic, an in-depth case study or clinical example and a presentation of core concepts that students must understand in order to make independent contributions.

More information about this series at <http://www.springer.com/series/10059>

Koteswara Rao Anne • Swarna Kuchibhotla
Hima Deepthi Vankayalapati

Acoustic Modeling for Emotion Recognition

 Springer

Koteswara Rao Anne
Dept of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada
India

Hima Deepthi Vankayalapati
Dept of Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha
Engineering College, Kanuru
Vijayawada
India

Swarna Kuchibhotla
Acharya Nagarjuna University
Andhra Pradesh
India

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISSN 2191-737X ISSN 2191-7388 (electronic)
SpringerBriefs in Speech Technology
ISBN 978-3-319-15529-6 ISBN 978-3-319-15530-2 (eBook)
DOI 10.1007/978-3-319-15530-2

Library of Congress Control Number: 2015934453

Springer Cham Heidelberg New York Dordrecht London
© The Author(s) - SpringerBriefs 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1	Introduction	1
1.1	Speech Signal Representation	1
1.2	Acoustic Signal Basics	2
1.3	Different Perspectives of Emotion	4
1.3.1	Physiology of Emotion	5
1.3.2	Computer Science Based Emotion	5
1.4	Applications of Speech Emotion Recognition	5
1.5	Book Organization	6
2	Emotion Recognition Using Prosodic Features	7
2.1	Introduction	7
2.2	Pre-Processing	8
2.2.1	Filtering	9
2.2.2	Framing	9
2.2.3	Windowing	10
2.3	Extraction of Prosodic Features	11
2.3.1	Zero Crossing Rate	11
2.3.2	Short Time Energy	12
2.3.3	Pitch	12
2.4	Importance of Prosodic Features	15
3	Emotion Recognition Using Spectral Features	17
3.1	Introduction	17
3.2	Extraction of Spectral Features	17
3.2.1	MelFrequency Cepstral Coefficients (MFCC)	17
3.2.2	Linear Predicticton Cepstral Coefficients (LPCC)	22
3.2.3	Formant Features	24
3.3	Importance of Spectral Features	24
4	Feature Fusion Techniques	27
4.1	Introduction	27
4.2	Multi Modal Feature Fusion	28

4.2.1	Adaptive and Non Adaptive Feature Fusion	28
4.2.2	Feature Level Feature Fusion	30
4.2.3	Decision Level Feature Fusion	30
5	Emotional Speech Corpora	33
5.1	Introduction	33
5.2	Overview of Emotional Speech Databases	33
5.3	Berlin Emotional Speech Database	34
5.4	Spanish Emotional Speech Database	34
5.5	Real Time Emotional Database of a Driver	35
6	Classification Models	37
6.1	Introduction	37
6.2	General Classification of Classification Models	37
6.2.1	Statistical Approach	38
6.2.2	Syntactic or Structural Approach	38
6.2.3	Template Matching	38
6.2.4	Neural Networks	38
6.3	Selected Classification Models	39
6.3.1	Linear Discriminant Analysis	40
6.3.2	Regularized Discriminant Analysis	44
6.3.3	Support Vector Machine	45
6.3.4	K-Nearest Neighbor	47
6.4	Distance Measures	48
6.4.1	Euclidian Distance	48
6.4.2	Standardized Euclidian Distance	49
6.4.3	Mahalanobis Distance	49
6.4.4	Cityblock Distance	49
6.4.5	Minkowski Distance	50
6.4.6	Cosine Distance	50
6.4.7	Hausdorff Distance	51
6.4.8	The Directed Hausdorff Distance	51
6.5	Summary	53
7	Comparative Analysis of Classifiers in Emotion Recognition	55
7.1	Introduction	55
7.2	Emotions Used in This Work	55
7.3	Analysis of Results with Each Emotion for Each Classifier	55
7.4	Analysis of Results with Confusion Matrices	56
7.5	Brief Overview of ROC Curves	58
7.5.1	Analysis of Results with ROC Curves	59
7.6	Summary	60
8	Summary and Conclusions	61
8.1	Summary of the Present Work	61

Contents	vii
8.2 Conclusion from the Present Work	62
8.3 Scope and Future Work	62
References	63