

Autor :

Prof. Dr.-Ing. Hans-Günter Hirsch

Lehrgebiet Nachrichtenübertragung, Lineare Systeme und
Netzwerke

Fachbereich Elektrotechnik und Informatik

Hochschule Niederrhein

Schriftenreihe des Fachbereichs Elektrotechnik und Informatik

herausgegeben von

Prof. Dr.-Ing. Hans Dieter Beims
Fachbereich Elektrotechnik und Informatik
Hochschule Niederrhein

Band 6/2008

Hans-Günter Hirsch

**Automatic Speech Recognition in
Adverse Acoustic Conditions**

Shaker Verlag
Aachen 2008

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Copyright Shaker Verlag 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8322-7226-5

ISSN 1610-9392

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

Most of today's speech recognition systems are based on a cepstral analysis scheme for representing short-term speech features. Furthermore, the statistical approach of modeling speech units by Hidden Markov Models (HMMs) and the application of the Viterbi algorithm are state-of-the-art for the pattern recognition in this field. These analysis and modeling techniques are presented as an introduction to the problem of recognizing speech in adverse acoustic conditions.

Analyzing the speech input in application scenarios of recognition systems the speech signal is affected by the acoustic environment. An overview about the combination of possible distortion effects is given. A simulation tool is presented that has been developed to allow the generation of distorted speech data for a lot of different acoustic environments. A new database is described that has been created with this tool. It is publicly available for comparative investigations on the recognition of distorted speech data.

After an introduction to the principal approaches for improving the recognition in adverse conditions a few details of a standardized method are presented that aims at the extraction of robust features. The alternative approach of adapting certain HMM parameters in adverse conditions is introduced by presenting details about a new method to adapt the HMM parameters in case of a hands-free speech input in rooms. Furthermore, it is shown that this technique can be combined with an adaptation to stationary background noise and unknown frequency characteristics. We demonstrate the efficiency of the new approaches by evaluating the results for recognizing artificially distorted data as well as speech data that have been recorded in noisy environments.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Structure of Speech Recognition Systems | 2 |
| 2.1 | MEL Frequency Cepstral Analysis | 3 |
| 2.2 | Modeling Speech Units as HMMs | 5 |
| 3 | Acoustic Scenarios during Speech Input | 9 |
| 3.1 | Simulation of the Acoustic Environment | 10 |
| 3.2 | Speech Database Aurora-5 | 14 |
| 3.2.1 | Test Sets | 14 |
| 3.2.2 | Training Sets | 17 |
| 3.2.3 | Recognition Experiments | 17 |
| 3.3 | Recognition Results for Different Distortion Effects | 18 |
| 4 | Improving the Recognition Performance in Adverse Conditions | 23 |
| 4.1 | Robust Feature Extraction Scheme | 25 |
| 4.2 | Adapting HMMs to Reverberation | 26 |
| 4.2.1 | Modeling the Influence of a Hands-free Speech Input | 26 |
| 4.2.2 | Adaptation of Static Parameters | 28 |
| 4.2.3 | Estimation of T_{60} | 32 |
| 4.3 | Adaptation of Delta Parameters | 33 |
| 4.4 | Restrictions in Case of a Connected Word Recognition | 37 |
| 4.5 | Adaptation of Triphone HMMs | 37 |
| 4.6 | Recognition Experiments on Hands-free Speech Input | 39 |
| 4.6.1 | Feature Extraction and Recognition Setup | 39 |
| 4.6.2 | Recognition with Whole Word HMMs | 41 |
| 4.6.3 | Recognition with Triphone HMMs | 45 |
| 4.7 | Combined Adaptation to All Distortion Effects | 46 |
| 4.7.1 | Estimation of Distortion Parameters | 47 |
| 4.7.2 | Combined Adaptation Scheme | 49 |
| 4.8 | Recognition Experiments on Hands-free Speech Input in Noisy Environments | 51 |
| 4.8.1 | Recognition Results for the Aurora-5 Database | 51 |
| 4.8.2 | Recognition Results for Noisy Recordings | 55 |
| 5 | Conclusions | 57 |
| | Glossary | 58 |

References**63**