**Statistical**

**Optimization**

**in**



**Speech**

**Recognition**

**Joseph Picone**

**Statistical Optimization**

**in Speech Recognition**

Joseph Picone

Modern speech recognition systems are testaments to advances in machine learning as they optimize every aspect of the system using statistically-based learning algorithms. Optimization has allowed this technology to achieve unprecedented levels of performance. This book presents a unified view of this optimization, focusing on such techniques as expectation-maximization, hidden Markov modeling and discriminative training. Theory is presented in a clear and straightforward manner, so readers can easily access and apply the material.

Joseph Picone has conducted research in the field of speech processing for more than 25 years in the industrial, academic and government sectors. He founded the Institute for Signal and Information Processing which is well-known for producing public domain speech and signal processing technology and providing innovative solutions over the Internet.

Statistical Optimization in
Speech Recognition

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Signal and Information Processing

J. Picone, *Fundamentals of Speech Recognition*, 1998

J. Picone, *et al*., *Pattern Recognition in Java*, 1998

J. Picone, A. Ganapathiraju and J. Hamaker, *Risk Minimization in Speech Recognition*, 2003

J. Picone, *Statistical Optimization in Speech Recognition*, 2009

J. Picone, *Perspectives on Engineering Management*, 2010

**Statistical Optimization in** Joseph Picone

**Speech Recognition**

The ISIP Press

City, State, USA

City, State, Country

© 2009 The Institute for Signal and Information Processing

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times New Roman using Microsoft Word 2007.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Picone, Joseph, 1957–

Statistical optimization in speech recognition / Joseph Picone.

p. cm. – (Language, speech, and communication)

Includes bibliographical references and index.

ISBN X-XXX-XXXXX-5 (alk. paper)

1. Automatic speech recognition—Statistical optimization. I Title.

II. Series.

TKXXXX.XXXXXX 2009 XX-XXXXX

XXX.X'XX—xxXX XXX

**DEDICATION**

To Mary Ann, Mary Frances, and Ann Marie, for their sacrifices that allowed me to spend many years away from home chasing dreams…

To Rafid, Scott, Rick, Paul, Julie, Aravind, Jon, Daniel, Lorena and many, many other ISIP members who contributed to this work and continue to motivate me to chase my dreams…

To Demetrius, Walt, Joe, George, Raja, Jack, Barb, Mark and many other colleagues who helped me turn some of those dreams into reality…

And to Jack for challenging me to dream about speech processing over a frozen cup of coffee in the middle of a cold Chicago winter many, many years ago…

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Contents

[Preface viii](#_Toc190485775)

[Chapter 1 Intro 10](#_Toc190485776)

[1.1. My First ASR 10](#_Toc190485777)

[Chapter 2 Features 10](#_Toc190485778)

[2.1. My Second ASR 10](#_Toc190485779)

[Chapter 3 Something Else 10](#_Toc190485780)

[Index 14](#_Toc190485781)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Preface

Optimization has played an enormously significant role in speech processing, particularly speech recognition. Virtually all successful speech technology today, including speech compression for cellular phones, speech synthesis for interactive voice response systems, and, speech recognition, employ some form of an optimization algorithm rooted in statistics to estimate parameters. I was first introduced to optimization theory as a college student in the late 1970’s, when a minimum mean square error approach to parameter estimation was used to model the speech signal as a digital filter. The technique, known as linear prediction, offered the promise of making digital speech processing a reality, because it was computationally efficient as well as effective. There was great hope that this optimization process could uncover hidden structure in the signal, a critical requirement for pattern recognition applications such as speech recognition.

More powerful techniques were beginning to emerge at this time, but they exceeded the computational resources available to most researchers. As high-speed computing became cost-effective over the next decade, the field quickly advanced by exploiting optimization principles in every aspect of the speech problem. Today, modern speech recognition systems use tens of thousands of hours of speech data to train millions of parameters to achieve unprecedented levels of performance.

Speech recognition has always been an excellent application for advanced statistical techniques. For many years, this research area was a vanguard for digital signal processing due to a combination of commercial and military interest in the technology, and properties of the application that make it a fertile ground for experimentation with new algorithms. In the past two decades, other signal processing fields have perhaps closed this gap and there has been more of a give and take on the assimilation of new research. But statistical optimization remains an important underpinning of all of these fields.

It is the goal of this book to provide some perspective on these optimization techniques. This book is not intended to be a comprehensive treatment of modern speech recognition technology, because the field has become too vast to make that possible. State of the art systems draw heavily on many fields including linguistics, signal processing, and natural language processing, and involve hundreds of steps to train and adapt high performance models. A comprehensive textbook can only give a cursory treatment of the basics and cannot do justice to the many practical issues that must be dealt with to achieve success.

Instead, this book focuses on a single theme, optimization, and discusses how this principle is applied to each major component of a speech recognition system. It is intended for those who are not familiar with speech recognition and can perhaps gain some knowledge that can be applied to other domains. It is also geared towards people new to the field who would like to gain some perspective on the evolution of the field. Optimization has been a recurring theme in the past 20 years. Significant amounts of research have focused on new parameter estimation techniques designed to minimize error rates.

Central to this book is the concept of hidden Markov modeling. This has been the driving force in speech recognition for the past 30 years. But it is the goal of this book to emphasize the optimization and robust parameter estimation aspects of this approach, not to focus on the details of how a state of the art speech recognition system is constructed using this approach.

Readers should be familiar with the basics of probability theory, information theory and pattern recognition. Whenever possible, we attempt to make the explanations and derivations self-contained so the book can serve as a useful reference book. We also include, where appropriate, discussions of practical issues that are essential to realizing an effective implementation of an algorithm. We attempt to keep mathematical notation clean and simple, and more importantly, consistent throughout the book. [Provide some specifics…]

The book is organized into three sections. First, the introductory portion of the book focuses on two topics: the basic elements of a speech recognition system and an overview of statistical optimization theory from a speech processing perspective. Second, we discuss a broad class of generative techniques that attempt to model variation in the signal by increasing the likelihood of the training data given a model. These techniques remain in use today though their origins date back to the mid-1980’s. The third section of the book discusses discriminative techniques that are designed to directly minimize recognition error rate. The last two sections are organized according to the basic system architecture discussed in the first section, and provide examples of how the techniques are applied to specific elements of a speech recognition system. It is our hope that this perspective, somewhat nontraditional for speech processing textbooks, will emphasize the important underlying principles of these approaches, and serve as a springboard to additional studies on more advanced topics.

This book evolved from many years of experience teaching speech processing courses at the graduate level extending back to the early 1980’s when our focus was more on dynamic programming and signal processing. One of the greatest influences on this book was the work done by a small, dedicated research group known as the Institute for Signal and Information Processing (ISIP), which can be accessed from the URL *http://www.ece.msstate.edu/research/isip*. It was here we developed one of the first open source speech recognition systems, funded by the National Science Foundation and the Department of Defense. We also developed a wealth of educational materials, all available from the ISIP web site, that have been instrumental in formulating clear and concise ways to explain various aspects of the speech recognition problem. Many students have contributed to these materials over the years, and all deserve thanks for their hard work, extreme dedication and excellent contributions. We were also fortunate to have many visitors to our laboratory who participated in our summer workshops, used our software, and gave us excellent feedback on our pedagogy. Their feedback was also instrumental in understanding how to convey this material to aspiring researchers.

The second significant influence on this book was the pioneering research my colleagues Aravind Ganapathiraju and Jon Hamaker performed as graduate students at ISIP. Aravind’s work on Support Vector Machines laid the groundwork for the unified view presented here. His in-depth analysis of learning machines laid the groundwork for many years of research in ISIP on statistical modeling. Jon Hamaker’s work on risk minimization similarly provided a valuable perspective on Bayesian methods and their role in the construction of learning machines. They started the tradition of Philosophical Fridays – a time when anyone could get up on a soapbox as long as they brought the bagels and coffee – and that was the genesis of our research in this area.

Special thanks go to [people to be named later…] for reviewing and commenting on this manuscript, and to many people outside of ISIP who accessed preliminary versions of this manuscript from the ISIP web site and provided excellent feedback. I should also express a note of thanks to Microsoft for providing Word 2007, which made it possible to prepare this manuscript in a relatively straightforward manner. It only took about half the amount of time it would have taken in Word 2003, and only twice the amount of time it would have taken in Adobe Framemaker ☺

Joseph Picone

*Crawford, Mississippi*

*February 2008*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Intro

This is a test…

## My First ASR

# Features

## My Second ASR

# Something Else

What is this book?

What is covered in the book?

This is the preface. I hope it is really good because it needs to be.

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting

Say something really interesting here… For example, speech recognition is really cool! Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… John Smith

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here… Say something really interesting here…

a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is

This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is

Jane Doe This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This This is a test.

This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. Feature extraction This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This This is a test.

This is a test. This is a test. This is a test. This is a test. This is a test. Fourier transform under feature extraction This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test.

This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is a test. This is This is a test. This is a test. This is a test. This

#

# Index

F

Feature Extraction, **3**

Fourier transform, 3

S

speech recognition, 1