The Algorithms Of Speech Recognition Programming And

Refereed postproceedings of the International Conference on Non-Linear Speech Processing, NOLISP 2005. The 30 revised full papers presented together with one keynote speech and 2 invited talks were carefully reviewed and selected from numerous submissions for inclusion in the book. The papers are organized in topical sections on speaker recognition, speech analysis, voice pathologies, speech recognition, speech enhancement, and applications.

This book introduces the theory, algorithms, and implementation techniques for efficient decoding in speech recognition mainly focusing on the Weighted Finite-State Transducer (WFST) approach. The decoding process for speech recognition is viewed as a search problem whose goal is to find a sequence of words that best matches an input speech signal. Since this process becomes computationally more expensive as the system vocabulary size increases, research has long been devoted to reducing the computational cost. Recently, the WFST approach has become an important state-of-the-art speech recognition technology, because it offers improved decoding speed with fewer recognition errors compared with conventional methods. However, it is not easy to understand all the algorithms used in this framework, and they are still in a black box for many people. In this book, we review the WFST approach and aim to provide comprehensive interpretations of WFST operations and decoding algorithms to help anyone who wants to understand, develop, and study WFST-based speech recognizers. We also mention recent advances in this framework and its applications to spoken language processing. Table of Contents: Introduction / Brief Overview of Speech Recognition / Introduction to Weighted Finite-State Transducers / Speech Recognition by Weighted Finite-State Transducers / Dynamic Decoders with On-the-fly WFST Operations / Summary and Perspective

Chapters in the first part of the book cover all the essential speech processing techniques for building robust, automatic speech recognition systems: the representation for speech signals and the methods for speech-features extraction, acoustic and language modeling, efficient algorithms for searching the hypothesis space, and multimodal approaches to speech recognition. The last part of the book is devoted to other speech processing applications that can use the information from automatic speech recognition for speaker identification and tracking, for prosody modeling in emotion-detection systems and in other speech processing applications that are able to operate in real-world environments, like mobile communication services and smart homes.

This book provides a cross-disciplinary reference to speech in mobile and pervasive environments Speech in Mobile and Pervasive Environments addresses the issues related to speech processing on resource-constrained mobile devices. These include speech recognition in noisy environments, specialised hardware for speech recognition and synthesis, the use of context to enhance recognition and user experience, and the emerging software standards required for interoperability. This book takes a multi-disciplinary look at these matters, while offering an insight into the opportunities and challenges of speech processing in mobile environs. In developing regions, speech-on-mobile is set to play a momentous role, socially and economically; the authors discuss how voice-based solutions and applications offer a compelling and natural solution in this setting. Key Features Provides a holistic overview of all speech technology related topics in the context of mobility Brings together the latest research in a logically connected way in a single volume Covers hardware, embedded recognition and synthesis, distributed speech recognition, software technologies, contextual interfaces Discusses multimodal dialogue systems and their evaluation Introduces speech in mobile and pervasive environments for developing regions This book provides a comprehensive overview for beginners and experts alike. It can be used as a textbook for advanced undergraduate and postgraduate students in electrical engineering and computer science. Students, practitioners or researchers in the areas of mobile computing, speech processing, voice applications, human-computer interfaces, and information and communication technologies will also find this reference insightful. For experts in the above domains, this book complements their strengths. In addition, the book will serve as a guide to practitioners working in telecom-related industries.

This E-book is a collection of articles that describe advances in speech recognition technology. Robustness in speech recognition refers to the need to maintain high speech recognition accuracy even when the quality of the input speech is degraded, or when the acoustical, articulate, or phonetic characteristics of speech in the training and testing environments differ. Obstacles to robust recognition include acoustical degradations produced by additive noise, the effects of linear filtering, nonlinearities in transduction or transmission, as well as impulsive interfering sources, and diminished accuracy caused by changes in articulation produced by the presence of high-intensity noise sources. Although progress over the past decade has been impressive, there are significant obstacles to overcome before speech recognition systems can reach their full potential. Automatic speech recognition (ASR) systems must be robust to all levels, so that they can handle background or channel noise, the occurrence on unfamiliar words, new accents, new users, or unanticipated inputs. They must exhibit more 'intelligence' and integrate speech with other modalities, deriving the user's intent by combining speech with facial expressions, eye movements, gestures, and other input features, and communicating back to the user through multimedia responses. Therefore, as speech recognition technology is transferred from the laboratory to the marketplace, robustness in recognition becomes increasingly significant. This E-book should be useful to computer engineers interested in recent developments in speech recognition technology.

Speech recognition technique has proven to be significantly beneficial in the domain of Artificial Intelligence. This book is based on speech processing and recognition. It consists of information provided by top researchers from Italy, Tunisia, India, Netherlands, Canada and Finland. Topics like speech recognition, noise cancellation, speech enhancement and emotion recognition have been described in this book. Important techniques like voice conversion and multi resolution spectral analysis have also been elucidated. The book consists of both original research works as well as surveys along with the applications of the technology in various scientific fields. The aim of this book is to serve as a good source of

knowledge for students and researchers related to this field.
Robust Automatic Speech Recognition: A Bridge to Practical Applications establishes a solid foundation for automatic speech recognition that is robust against acoustic environmental distortion. It provides a thorough overview of classical and modern noise-and reverberation robust techniques that have been developed over the past thirty years, with an emphasis on practical methods that have been proven to be successful and which are likely to be further developed for future applications. The strengths and weaknesses of robustness-enhancing speech recognition techniques are carefully analyzed. The book covers noise-robust techniques designed for acoustic models which are based on both Gaussian mixture models and deep neural networks. In addition, a guide to selecting the best methods for practical applications is provided. The reader will: Gain a unified, deep and systematic understanding of the state-of-the-art technologies for robust speech recognition Learn the links and relationship between alternative technologies for robust speech recognition Be able to use the technology analysis and categorization detailed in the book to guide future technology development Be able to develop new noise-robust methods in the current era of deep learning for acoustic modeling in speech recognition The first book that provides a comprehensive review on noise and reverberation robust speech recognition methods in the era of deep neural networks Connects robust speech recognition techniques to machine learning paradigms with rigorous mathematical treatment Provides elegant and structural ways to categorize and analyze noise-robust speech recognition techniques Written by leading researchers who have been actively working on the subject matter in both industrial and academic organizations for many years
Dynamic Speech Models provides a comprehensive overview of mathematical models of speech dynamics and addresses the following issues:" How do we make sense of the complex speech process in terms of its functional role of speech communication?" How do we quantify the special role of speech timing?" How do the dynamics relate to the variability of speech which has often been said to seriously hamper automatic speech recognition?" How do we put the dynamic process of speech into a quantitative form to enable detailed analyses?" How can we incorporate the knowledge of speech dynamics into computerized speech analysis and recognition algorithms?The answers to all these questions require building and applying computational models for the dynamic speech process. Such scientific studies help understand why humans speak as they do and how humans exploit redundancy and variability by way of multi-tiered dynamic processes to enhance the efficiency and effectiveness of speech. Second, advancement of human language technology, especially in automatic recognition of human speech is expected to benefit from comprehensive computational modeling of speech dynamics.The limitations of current speech recognition technology are serious and are well known. A commonly acknowledged and frequently discussed weakness of the statistical model underlying current speech recognition technology is the lack of adequate dynamic modeling schemes to provide correlation structure across the temporal speech observation sequence. Dynamic speech modeling may serve as an ultimate solution to this problem.
Este documento tiene por objetivo recoger la información relativa al proyecto sobre la creación de un algoritmo para Start-and-End point detection de una señal pregrabada. La intención inicial del desarrollo de este algoritmo es que pueda ser utilizado en la entrada de una aplicación de reconocimiento de voz. En términos generales, el resultado de este trabajo es un algoritmo que puede detectar el comienzo y el fin de una señal previamente grabada basado en un algoritmo de detección de la actividad de la voz previamente desarrollado por la Czech Technical University, Faculty of Electrical Engineering. Hay dos temas principales de estudio en este proyecto: detección de la actividad de la voz (VAD algorithm) y determinar el punto de inicio y fin de la señal (Start-and-End point detection). El primer paso para la construcción del algoritmo final es ser capaz de identificar la actividad de la voz en una señal mediante el VAD algorithm para después ser capaz de detectar el inicio y final de la actividad de la voz y descartar los silencios de la señal mediante el Startand- End point detection algorithm. Con el fin de demostrar el modo de funcionamiento de dicho algoritmo se ha creado una aplicación en MATLAB que permite ver gráficamente una señal previamente grabada y posteriormente su punto inicial y final después de aplicar los algoritmos. Por último, para proporcionar resultados más gráficos y dar al proyecto un valor añadido y con vistas a convertirse en una futura aplicación posible se ha añadido el reconocimiento de dígitos basado en de un algoritmo DTW (Dinamic Time Warping). English: The objective of the project is the creation of an algorithm for Start-and-End point detection of a pre-recorded signal. The initial reason for developing this algorithm is so it can be used at the input of a voice recognition application. Overall, the result of this work is an algorithm that can detect the beginning and end of a previously recorded signal based on a detection algorithm of the voice activity previously developed by the Czech Technical University, Faculty of Electrical Engineering. Two main issues are studied in this project: Detecting the Voice Activity (VAD algorithm) and determining the start and end point of the signal (Start-and-End point detection). To demonstrate the mode of operation of the algorithm, I have created an application in MATLAB to show graphically the process for a previously recorded signal and then the start and end points after applying the algorithms. Finally, to provide better graphic performance and provide added value to the project, I have added a digit recognition algorithm based on a DTW (Dynamic Time Warping).
This textbook explains Deep Learning Architecture, with applications to various NLP Tasks, including Document Classification, Machine Translation, Language Modeling, and Speech Recognition. With the widespread adoption of deep learning, natural language processing (NLP),and speech applications in many areas (including Finance, Healthcare, and Government) there is a growing need for one comprehensive resource that maps deep learning techniques to NLP and speech and provides insights into using the tools and libraries for real-world applications. Deep Learning for NLP and Speech Recognition explains recent deep learning methods applicable to NLP and speech, provides state-of-the-art approaches, and offers real-world case studies with code to provide hands-on experience. Many books focus on deep learning theory or deep learning for NLP-specific tasks while others are cookbooks for tools and libraries, but the constant flux of new algorithms, tools, frameworks, and libraries in a rapidly evolving landscape means that there are few

available texts that offer the material in this book. The book is organized into three parts, aligning to different groups of readers and their expertise. The three parts are: Machine Learning, NLP, and Speech Introduction The first part has three chapters that introduce readers to the fields of NLP, speech recognition, deep learning and machine learning with basic theory and hands-on case studies using Python-based tools and libraries. Deep Learning Basics The five chapters in the second part introduce deep learning and various topics that are crucial for speech and text processing, including word embeddings, convolutional neural networks, recurrent neural networks and speech recognition basics. Theory, practical tips, state-of-the-art methods, experimentations and analysis in using the methods discussed in theory on real-world tasks. Advanced Deep Learning Techniques for Text and Speech The third part has five chapters that discuss the latest and cutting-edge research in the areas of deep learning that intersect with NLP and speech. Topics including attention mechanisms, memory augmented networks, transfer learning, multi-task learning, domain adaptation, reinforcement learning, and end-to-end deep learning for speech recognition are covered using case studies.

This book discusses large margin and kernel methods for speech and speaker recognition Speech and Speaker Recognition: Large Margin and Kernel Methods is a collation of research in the recent advances in large margin and kernel methods, as applied to the field of speech and speaker recognition. It presents theoretical and practical foundations of these methods, from support vector machines to large margin methods for structured learning. It also provides examples of large margin based acoustic modelling for continuous speech recognizers, where the grounds for practical large margin sequence learning are set. Large margin methods for discriminative language modelling and text independent speaker verification are also addressed in this book. Key Features: Provides an up-to-date snapshot of the current state of research in this field Covers important aspects of extending the binary support vector machine to speech and speaker recognition applications Discusses large margin and kernel method algorithms for sequence prediction required for acoustic modeling Reviews past and present work on discriminative training of language models, and describes different large margin algorithms for the application of part-of-speech tagging Surveys recent work on the use of kernel approaches to text-independent speaker verification, and introduces the main concepts and algorithms Surveys recent work on kernel approaches to learning a similarity matrix from data This book will be of interest to researchers, practitioners, engineers, and scientists in speech processing and machine learning fields.

We have formulated new algorithms for joint evaluation of the likelihood of multiple speech patterns, using the standard HMM framework. This was possible through the judicious use of the basic DTW algorithm extended to multiple patterns. We also showed that this joint formulation is useful in selective training of HMMs, in the context of burst noise or mispronunciation among training patterns. Although these algorithms are evaluated in the context of IWR under burst noise conditions, the formulation and algorithm can be useful in different contexts, such as connected word recognition (CWR) or continuous speech recognition (CSR). In spoken dialog systems, if the confidence level of the test speech is low, the system can ask the user to repeat the pattern. However, in the continuous speech recognition case, a user cannot be expected to repeat a sentence/s exactly. But still the proposed methods can be used. Here is one scenario. For booking a railway ticket, the user says, "I want a ticket from Bangalore to Aluva". The recognition system asks the user, "Could you please repeat from which station would you like to start?". The user repeats the word "Bangalore". So this word "Bangalore" can be jointly recognized with the word "Bangalore" from the first sentence to improve speech recognition performance. One of the limitations of the new formulation is when the whole pattern is noisy, i.e., when the noise is continuous not bursty; the proposed algorithms don't work well. Also, for the present, we have not addressed the issue of computational complexity, which is high in the present implementations. Efficient variations of these algorithms have to be explained for real-time or large scale CSR applications. Finally we conclude that jointly evaluating multiple speech patterns is very useful for speech training and recognition and it would greatly aid in solving the automatic speech recognition problem. We hope that our work will show a new direction of research in this area.

Speech Processing has rapidly emerged as one of the most widespread and well-understood application areas in the broader discipline of Digital Signal Processing. Besides the telecommunications applications that have hitherto been the largest users of speech processing algorithms, several non-traditional embedded processor applications are enhancing their functionality and user interfaces by utilizing various aspects of speech processing. "Speech Processing in Embedded Systems" describes several areas of speech processing, and the various algorithms and industry standards that address each of these areas. The topics covered include different types of Speech Compression, Echo Cancellation, Noise Suppression, Speech Recognition and Speech Synthesis. In addition this book explores various issues and considerations related to efficient implementation of these algorithms on real-time embedded systems, including the role played by processor CPU and peripheral functionality.

Automatic Speech Recognition (ASR) is the enabling technology for hands-free dictation and voice-triggered computer menus. It is becoming increasingly prevalent in environments such as private telephone exchanges and real-time information services. Speech Recognition introduces the principles of ASR systems, including the theory and implementation issues behind multi-speaker continuous speech recognition. Focusing on the algorithms employed in commercial and laboratory systems, the treatment enables the reader to devise practical solutions for ASR system problems. It addresses in detail C++ programming techniques used to develop ASR applications, thus offering skills that will prove useful in any large C++ based software project. Possible extensions of the well-established ASR technology are highlighted, based on "Hidden Markov Models" applied to fields such as modelling and prediction of econometric series. Features include: * Accompanying website containing all C++ source code of a complete laboratory multi-speaker continuous-speech ASR system (e.g. Initialisation, Training, Recognition, Evaluation, etc.) www.wiley.com/go/becchetti_speech * Detailed theoretical, mathematical and technical explanations of ASR * A practical account of the functioning of ASR A crucial source of information for researchers, developers and project managers

involved with ASR systems, Speech Recognition is also structured for use by students of digital signal processing, speech recognition and C++ programming techniques.

This book on Speech Processing consists of seven chapters written by eminent researchers from Italy, Canada, India, Tunisia, Finland and The Netherlands. The chapters covers important fields in speech processing such as speech enhancement, noise cancellation, multi resolution spectral analysis, voice conversion, speech recognition and emotion recognition from speech. The chapters contain both survey and original research materials in addition to applications. This book will be useful to graduate students, researchers and practicing engineers working in speech processing.

It is with great pleasure that I present this third volume of the series "Advanced Applications in Pattern Recognition." It represents the summary of many man- (and woman-) years of effort in the field of speech recognition by tne author's former team at the University of Turin. It combines the best results in fuzzy-set theory and artificial intelligence to point the way to definitive solutions to the speech-recognition problem. It is my hope that it will become a classic work in this field. I take this opportunity to extend my thanks and appreciation to Sy Marchand, Plenum's Senior Editor responsible for overseeing this series, and to Susan Lee and Jo Winton, who had the monumental task of preparing the camera-ready master sheets for publication. Morton Nadler General Editor vii PREFACE Si parva licet componere magnis Virgil, Georgics, 4,176 (37-30 B.C.) The work reported in this book results from years of research oriented toward the goal of making an experimental model capable of understanding spoken sentences of a natural language. This is, of course, a modest attempt compared to the complexity of the functions performed by the human brain. A method is introduced for conce1v1ng modules performing perceptual tasks and for combining them in a speech understanding system.

Current Automatic Speech Recognition devices attempt to solve the connected word recognition problem by assuming that an unknown phrase is the output of a sequence of statistical word-models. Typically, these models are constructed using examples of words spoken in isolation; however, the acoustic patterns corresponding to words as they occur in fluent speech are quite different from those representing the same words spoken in isolation, and so the use in speech recognizers of models based on isolated utterances severely limits the performance of such devices. A method of extracting training utterances from fluent speech and constructing Hidden Markov Models (HMMs) from these templates, known as Embedded Training, is investigated here, in conjunction with a two-level algorithm for connected word recognition. The effects on recognition performance of various HMM training procedures are discussed, and experimental results are presented.

This book presents the revised tutorial lectures given at the International Summer School on Nonlinear Speech Processing-Algorithms and Analysis held in Vietri sul Mare, Salerno, Italy in September 2004. The 14 revised tutorial lectures by leading international researchers are organized in topical sections on dealing with nonlinearities in speech signals, acoustic-to-articulatory modeling of speech phenomena, data driven and speech processing algorithms, and algorithms and models based on speech perception mechanisms. Besides the tutorial lectures, 15 revised reviewed papers are included presenting original research results on task oriented speech applications.

The goal of speech recognition is to find the most probable word given the acoustic evidence, i.e. a string of VQ codes or acoustic features. Speech recognition algorithms typically take advantage of the fact that the probability of a word, given a sequence of VQ codes, can be calculated.

Comparison of Algorithms for Speech RecognitionNonlinear Analyses and Algorithms for Speech ProcessingInternational Conference on Non-Linear Speech Processing, NOLISP 2005, Barcelona, Spain, April 19-22, 2005, Revised Selected PapersSpringer

This chapter presents a family of feature compensation algorithms for noise robust speech recognition that use stereo data. The basic idea of the proposed algorithms is to stack the features of the clean and noisy channels to form a new augmented space, and to train.

Speech recognition algorithms were analyzed using normal and G-stressed speech as an input. Speech samples were recorded in centrifuge tests at the Air Force Medical Research Lab, Wright-Patterson AFB, Ohio. All speech was recorded using the MBU-12/P face mask. The algorithms studied are phoneme-based feature extractors which feed a recognition algorithm based on fuzzy set theory. Three feature extraction algorithm options were analyzed. One option used a phoneme length of 40 ms and the other options used a length of 8 ms. The recognition results for all three options using normal speech are above 90%, but the 40ms phoneme length give higher raw scores. For G-stressed speech the 40 ms phoneme length scored greater than 90% while the 8ms phoneme length options scored less than 60%. (Author).

This dissertation focuses on robust signal processing algorithms for birdsongs and speech signals. Automatic phrase or syllable detection systems of bird sounds are useful in several applications. However, bird-phrase detection is challenging due to segmentation error, duration variability, limited training data, and background noise. Two spectrograms with identical class labels may look different due to time misalignment and frequency variation. In real recording environments such as in a forest, the data can be corrupted by background interference, such as rain, wind, other animals or even other birds vocalizing. A noise-robust classifier needs to handle such conditions. Similarly, Automatic Speech Recognition (ASR) works well in quiet environments, but a large degradation in performance is observed when the speech signal is corrupted by background noise. The ASR performance would benefit from robust representations of speech signals and from robust recognition systems. The first topic of this dissertation focuses on an automatic birdsong-phrase recognition system that is robust to limited training data, class variability, and noise. The algorithm comprises a noise-robust Dynamic-Time-Warping (DTW)- based segmentation and a discriminative classifier for outlier rejection. The algorithm utilizes DTW and prominent (high energy) time-frequency regions of training spectrograms to derive a reliable noise-robust template for each phrase class. The resulting template is then used for segmenting continuous recordings to obtain segment candidates whose spectrogram amplitudes in the prominent regions are used as features to a Support Vector Machine (SVM). In addition, we present a novel approach to training HMMs with extremely limited data. First, the algorithm learns the Global Gaussian Mixture Models (GMMs) for all training phrases available. GMM parameters are then used to initialize state parameters of each individual model. The number of states and the mixture components for each state are determined by the

acoustic variation of each phrase type. The (high-energy) time-frequency prominent regions are used to compute the state emitting probability to increase noise-robustness. The second topic of the dissertation deals with noise-robust processing for automatic speech recognition. We also propose a new pitch-based spectral enhancement algorithm based on voiced frames for speech analysis and noise-robust speech processing. The proposed algorithm determines a time-warping function (TWF) and the speaker's pitch with high precision, simultaneously. This technique reduces the smearing effect in between harmonics when the fundamental frequency is not constant within the analysis window. To do so, we propose a metric called the harmonic residual which measures the difference between the actual spectrum and the resynthesized spectrum derived from the linear model of speech production with various combinations of TWF and high-precision pitch values as parameters. The TWF and pitch pair that yields the minimum harmonic residual is selected and the enhanced spectrum is obtained accordingly. We show how this new representation can be also used for automatic speech recognition by proposing a robust spectral representation derived from harmonic amplitude interpolation.

Remarkable progress is being made in spoken language processing, but many powerful techniques have remained hidden in conference proceedings and academic papers, inaccessible to most practitioners. In this book, the leaders of the Speech Technology Group at Microsoft Research share these advances -- presenting not just the latest theory, but practical techniques for building commercially viable products.KEY TOPICS: Spoken Language Processing draws upon the latest advances and techniques from multiple fields: acoustics, phonology, phonetics, linguistics, semantics, pragmatics, computer science, electrical engineering, mathematics, syntax, psychology, and beyond. The book begins by presenting essential background on speech production and perception, probability and information theory, and pattern recognition. The authors demonstrate how to extract useful information from the speech signal; then present a variety of contemporary speech recognition techniques, including hidden Markov models, acoustic and language modeling, and techniques for improving resistance to environmental noise. Coverage includes decoders, search algorithms, large vocabulary speech recognition techniques, text-to-speech, spoken language dialog management, user interfaces, and interaction with non-speech interface modalities. The authors also present detailed case studies based on Microsoft's advanced prototypes, including the Whisper speech recognizer, Whistler text-to-speech system, and MiPad handheld computer.MARKET: For anyone involved with planning, designing, building, or purchasing spoken language technology.