# Automatic Speech Recognition On Mobile Devices And Over Communication Networks Advances In Computer Vision And Pattern Recognition

This is likewise one of the factors by obtaining the soft documents of this **Automatic Speech Recognition On Mobile Devices And Over Communication Networks Advances In Computer Vision And Pattern Recognition** by online. You might not require more epoch to spend to go to the books inauguration as with ease as search for them. In some cases, you likewise complete not discover the revelation Automatic Speech Recognition On Mobile Devices And Over Communication Networks Advances In Computer Vision And Pattern Recognition that you are looking for. It will totally squander the time.

However below, behind you visit this web page, it will be hence agreed simple to get as without difficulty as download lead Automatic Speech Recognition On Mobile Devices And Over Communication Networks Advances In Computer Vision And Pattern Recognition

It will not recognize many time as we tell before. You can attain it though show something else at house and even in your workplace. therefore easy! So, are you question? Just exercise just what we have the funds for below as skillfully as evaluation **Automatic Speech Recognition On Mobile Devices And Over**

**Communication Networks Advances In Computer Vision And Pattern Recognition** what you similar to to read!

Distant Speech Recognition Springer

Automatic Speech Recognition (ASR) is undoubtedly one of the most important and interesting applications in the cutting-edge era of Deep-learning deployment, especially in the mobile segment. Fast and accurate ASR comes at a high energy cost, requiring huge memory storage and computational power, which is not affordable for the tiny power budget of mobile devices. Hardware acceleration can reduce power consumption of ASR systems as well as reducing its memory pressure, while delivering high-performance. In this thesis, we present a customized accelerator for large-vocabulary, speaker-independent, continuous speech recognition. A state-of-the-art ASR system consists of two major components: acoustic-scoring using DNN and speech-graph decoding using Viterbi search. As the first step, we focus on the Viterbi search algorithm, that represents the main bottleneck in the ASR system. The accelerator includes some innovative techniques to improve the memory subsystem, which is the main bottleneck for performance and power, such as a prefetching scheme and a novel bandwidth saving technique tailored to the needs of ASR. Furthermore, as the speech graph is vast taking more than 1-Gigabyte memory space, we propose to change its representation by partitioning it into several sub-graphs and perform an on-the-fly composition during the Viterbi run-time. This approach together with some simple yet efficient compression techniques result in 31x memory footprint reduction, providing 155x real-time speedup and orders of magnitude power and energy saving compared to CPUs and GPUs. In the next step, we propose a novel hardware-based ASR system that effectively integrates a DNN accelerator for the pruned/quantized models with the Viterbi accelerator. We show that, when either pruning or quantizing the DNN model used for acoustic scoring, ASR accuracy is maintained but the execution time of the ASR system is increased by 33%. Although pruning and quantization improves the efficiency of the DNN, they result in a huge increase of activity in the Viterbi search since the output scores of the pruned model are less reliable. In order to avoid the aforementioned increase in Viterbi search workload, our system loosely selects the N-best hypotheses at every time step, exploring only the N most likely paths. Our final solution manages to efficiently combine both DNN and Viterbi accelerators using all their optimizations, delivering 222x real-time ASR with a small power budget of 1.26 Watt, small memory footprint of 41 MB, and a peak memory bandwidth of 381 MB/s, being amenable for low-power mobile platforms.

Mobile Multimedia Processing

Allied Publishers

Over the past few years, speech recognition technology performance on tasks ranging from isolated digit recognition to conversational speech has dramatically improved. Performance on limited recognition tasks in noise-free environments is comparable to that achieved by human transcribers. This advancement in automatic speech recognition technology along with an increase in the compute power of mobile devices, standardization of communication protocols, and the explosion in the popularity of the mobile devices, has created an interest in flexible voice interfaces for mobile devices. However, speech recognition performance degrades dramatically in mobile environments which are inherently noisy. In the recent past, a great amount of effort has been spent on the development of front ends based on advanced noise robust approaches. The primary objective of this thesis was to analyze the performance of two advanced front ends, referred to as the QIO and MFA front ends, on a speech recognition task based on the Wall Street Journal database. Though the advanced front ends are shown to achieve a significant improvement over an industry-standard baseline front end, this improvement is not operationally significant. Further, we show that the results of this evaluation were not significantly impacted by suboptimal recognition system parameter settings. Without any front end-specific tuning, the MFA front end outperforms the QIO front end by 9.6% relative. With tuning, the relative performance gap increases to 15.8%. Finally, we also show that mismatched microphone and additive noise evaluation conditions resulted in a significant degradation in performance for both front ends.

*Mobile Speech and Advanced Natural Language Solutions*
Springer Science & Business Media

Abstract: "Automatic speech recognition (ASR) is a computerized speech-to-text process, in which speech is usually recorded with acoustical microphones by capturing air pressure changes. This kind of air-transmitted speech signal is prone to two kinds of problems related to noise robustness and applicability. The former means the mixing of speech signal and ambient noise usually deteriorates ASR performance. The latter means speech could be overheard easily on the air-transmission channel, and this often results in privacy loss or annoyance to other people. This thesis research solves these two problems by using channels that contact the human body without air transmission, i.e., by vibrocervigraphic and electromyographic methods. The vibrocervigraphic (VCG) method measures the throat vibration with a ceramic piezoelectric transducer contact to the skin on the neck, and the electromyographic (EMG) method measures the muscular electric potential with a set of electrodes attached to the skin where the articulatory muscles underlie. The VCG and EMG methods are inherently more robust to ambient noise, and they make it possible to recognize whispered and silent speech to improve applicability. The major contribution of this dissertation includes feature design and adaptation for optimizing features, acoustic model adaptation for adapting traditional acoustic models onto different feature spaces, and articulatory feature classification for incorporating articulatory information to improve recognition. For VCG ASR, the combination of feature transformation methods and maximum a posteriori adaptation improves the recognition accuracy even with a very small data set. On top of that, additive performance gain is achieved by applying maximum likelihood linear regression and feature space adaptation with different data granularities in order to adapt to channel variations as well as to speaker variations. For EMG ASR, we propose the Concise EMG feature that extracts representative EMG characteristics. It improves the recognition accuracy and advances the EMG ASR research from isolated word recognition to phone-based continuous speech recognition. Articulatory features are studied in both VCG and EMG ASR to analyze the systems and improve recognition accuracy. These techniques are demonstrated to be effective on both experimental evaluations and prototype applications."

*Multilingual Phone Recognition in Indian Languages* John Wiley & Sons

Automatic speech recognition and speaker recognition have a lot of applications in personal identification, access control and in the new man-machine-interface paradigm. The existing applications in voice-activated embedded systems solve the problem of recognition of the spoken words only or the problem of recognition of a speaker through the words uttered only. The goal of this project, therefore, is the development of a robust algorithm for both speech recognition and speaker verification. An example of a target application of this work is speech dialing of mobile phones with a speaker verification front-end in order to effect access control. In view of the memory and computational constraints of embedded systems, the dynamic time warping algorithm is used. This project only considers isolated spoken digits. The developed algorithm is coded in C language and can be ported to firmware for Arabic numeral digit recognition with a speaker verification front end for an embedded system like mobile phones. The system produced a FAR of 13.33% and a FRR of 24.3% for a total of 70 true claims and 30 false claims. It also had a word accuracy of 96.7%.

**Automatic Speech Recognition** Springer Science & Business Media

The advances in computing and networking have sparked an enormous interest in deploying automatic speech recognition on mobile devices and over communication networks. This book brings together academic researchers and industrial practitioners to address the issues in this emerging realm and presents the reader with a comprehensive introduction to the subject of speech recognition in devices and networks. It covers network, distributed and embedded speech recognition systems.

IPhone BoD – Books on Demand

Foreword Looking back the past 30 years. we have seen steady progress made in the area of speech science and technology. I still remember the excitement in the late seventies when Texas Instruments came up with a toy named "Speak-and-Spell" which was based on a VLSI chip containing the state-of-the-art linear prediction synthesizer. This caused a speech technology fever among the electronics industry. Particularly. applications of automatic speech recognition were rigorously attempt ed by many companies. some of which were start-ups founded just for this purpose. Unfortunately. it did not take long before they realized that automatic speech rec ognition technology was not mature enough to satisfy the need of customers. The fever gradually faded away. In the meantime. constant efforts have been made by many researchers and engi neers to improve the automatic speech recognition technology. Hardware capabilities have advanced impressively since that time. In the past few years. we have been witnessing and experiencing the advent of the "Information Revolution." What might be called the second surge of interest to com mercialize speech technology as a natural interface for man-machine communication began in much better shape than the first one. With computers much more powerful and faster. many applications look realistic this time. However. there are still tremendous practical issues to be overcome in order for speech to be truly the most natural interface between humans and machines.

**Research on Speech Communication and Automatic Speech Recognition** John Wiley & Sons

Speech recognition technology is being increasingly employed in human-machine interfaces. A remaining problem however is the robustness of this technology to non-native accents, which still cause considerable difficulties for current systems. In this book, methods to overcome this problem are described. A speaker adaptation algorithm that is capable of adapting to the current speaker with just a few words of speaker-specific data based on the MLLR principle is developed and combined with confidence measures that focus on phone durations as well as on acoustic features. Furthermore, a specific pronunciation modelling technique that allows the automatic derivation of non-native pronunciations without using non-native data is described and combined with the previous techniques to produce a robust adaptation to non-native accents in an automatic speech recognition system.

**Robustness in Automatic Speech Recognition** Springer The portable device and mobile phone market has witnessed rapid growth in the last few years with the emergence of several revolutionary products such as mobile TV, converging iPhone and digital cameras that combine music, phone and video functionalities into one device. The proliferation of this market has further bene?ted from the competition in software and applications for smart phones such as Google ' s Android operating system and Apple ' s iPhone App-Store, stimulating tens of thousands of mobile applications that are made ava- able by individual and enterprise developers. Whereas the mobile device has become ubiquitous in people ' s daily life not only as a cellular phone but also as a media player, a mobile computing device, and a personal assistant, it is p- ticularly important to address challenges timely in applying advanced pattern recognition, signal, information and multimedia processing techniques, and new emerging networking technologies to such mobile systems. The primary objective of this book is to foster interdisciplinary discussions and research in mobile multimedia processing techniques, applications and s- tems, as well as to provide stimulus to researchers on pushing the frontier of emerging new technologies and applications. One attempt on such discussions was the organization of the First Int- national Workshop of Mobile Multimedia Processing (WMMP 2008), held in Tampa, Florida, USA, on December 7, 2008. About 30 papers were submitted from10countriesacrosstheUSA,Asia andEurope.

*Soft Computing*

Springer Science & Business Media

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.

*Robust Adaptation to Non-Native Accents in Automatic Speech Recognition* John Wiley & Sons

Speech Recognition has a long history of being one of the difficult problems in Artificial Intelligence and Computer Science. As one goes from problem solving tasks such as puzzles and chess to perceptual tasks such as speech and vision, the problem characteristics change dramatically: knowledge poor to knowledge rich; low data rates to high data rates; slow response time (minutes to hours) to instantaneous response time. These characteristics taken together increase the computational complexity of the problem by several orders of magnitude. Further, speech provides a challenging task domain which embodies many of the requirements of intelligent behavior: operate in real time; exploit vast amounts of knowledge, tolerate errorful, unexpected unknown input; use symbols and abstractions; communicate in natural language and learn from the environment. Voice input to computers offers a number of advantages. It provides a natural, fast, hands free, eyes free, location free input medium. However, there are many as yet unsolved problems that prevent routine use of speech as an input device by non-experts. These include cost, real time response, speaker independence, robustness to variations such as noise, microphone, speech rate and loudness, and the ability to handle non-grammatical speech. Satisfactory solutions to each of these problems can be expected within the next decade. Recognition of unrestricted spontaneous continuous speech appears unsolvable at present. However, by the addition of simple constraints, such as clarification dialog to resolve ambiguity, we believe it will be possible to develop systems capable of accepting very large vocabulary continuous speechdictation.

*Automatic Speech Recognition on Mobile Devices and over Communication Networks* Academic Press

This Edited Volume gathers a selection of refereed and revised papers originally presented at the Third International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS ' 17), held on September 13 – 16, 2017 in Manipal, India. The papers offer stimulating insights into biometrics, digital watermarking, recognition systems, image and video processing, signal and speech processing, pattern recognition, machine learning and knowledge-based systems. Taken together, they offer a valuable resource for all researchers and scientists engaged in the various fields of signal processing and related areas.

**Speech Recognition Over Digital Channels** Academic Press

The iPhone may be the world's coolest computer, but it's still a computer, with all of the complexities. iPhone: The Missing Manual is a illustrated guide to the tips, shortcuts, and workarounds that will turn you, too, into an iPhone master. This updated guide shows you everything you need to know about the new features and user interface of iOS 9 for the iPhone. This easy-to-use book will help you accomplish everything from web browsing to watching videos so you can get the most out of your iPhone.

Intelligent Speech Signal Processing Springer Science & Business Media

Learning vocabulary using

ASR is a unique concept since it demands language learners to speak up in order to find word definitions. With the advancement of technology, it has become possible to incorporate ASR to computers, mobile phones, and other smart digital devices to learn L2 vocabulary. Two groups of adult, ESOL learners were given a vocabulary activity which had to be completed in class by using Google and Apple based ASR assistants in their mobile phones. Then, they were given a brief questionnaire inquiring about their perceptions on using ASR in the classroom. After the study, the teachers of the two classes were also asked about their perceptions about using ASR in the classroom. The data gathered from students ' questionnaire was analyzed under three qualitative evaluation criteria. Finally, the students ' perceptions about ASR were compared with teachers ' perceptions. While students cited many advantages of using ASR in the classroom to learn vocabulary, overall they claimed that mispronunciation of words and verbal commands became a hindrance in learning L2 vocabulary via ASR. On the contrary, the teachers thought that ASR helped student to improve their pronunciation. Thus, there was also a mismatch between students ' perceptions and teachers ' perceptions.--

**Performance Analysis of Advanced Front Ends on the Aurora Large Vocabulary Evaluation** Springer

This book focuses primarily on speech recognition and the related tasks such as speech enhancement and modeling. This book comprises 3 sections and thirteen chapters written by eminent researchers from USA, Brazil, Australia, Saudi Arabia, Japan, Ireland, Taiwan, Mexico, Slovakia and India. Section 1 on speech recognition consists of seven chapters. Sections 2 and 3 on speech enhancement and speech modeling have three chapters each respectively to supplement section 1. We sincerely believe that thorough reading of these thirteen chapters will provide comprehensive knowledge on modern speech recognition approaches to the readers.

**A Multi-band Approach to Automatic Speech Recognition** Springer Automatic speech recognition suffers from a lack of robustness with respect to noise, reverberation and interfering speech. The growing field of speech recognition in the presence of missing or uncertain input data seeks to ameliorate those problems by using not only a preprocessed speech signal but also an estimate of its reliability to selectively focus on those segments and features that are most reliable for recognition. This book presents the state of the art in recognition in the presence of uncertainty, offering examples that utilize uncertainty information for noise robustness, reverberation robustness, simultaneous recognition of multiple speech signals, and audiovisual speech recognition. The book is appropriate for scientists and researchers in the field of speech recognition who will find an overview of the state of the art in robust speech recognition, professionals working in speech recognition who will find strategies for improving recognition results in various conditions of mismatch, and lecturers of advanced courses on speech processing or speech recognition who will find a reference and a comprehensive introduction to the field. The book assumes an understanding of the fundamentals of speech recognition using Hidden Markov Models.

*Language Modeling for Automatic Speech Recognition of Inflective Languages* Springer Science & Business Media

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.

Robust Speech Recognition of Uncertain or Missing Data

IntechOpen

The book presents current research and developments in multilingual speech recognition. The author presents a Multilingual Phone Recognition System (Multi-PRS), developed using a common multilingual phone-set derived from the International Phonetic Alphabets (IPA) based transcription of six Indian languages - Kannada, Telugu, Bengali, Odia, Urdu, and Assamese. The author shows how the performance of Multi-PRS can be improved using tandem features. The book compares Monolingual Phone Recognition Systems (Mono-PRS) versus Multi-PRS and baseline versus tandem system. Methods are proposed to predict Articulatory Features (AFs) from spectral features using Deep Neural Networks (DNN). Multitask learning is explored to improve the prediction accuracy of AFs. Then, the AFs are explored to improve the performance of Multi-PRS using lattice rescoring method of combination and tandem method of combination. The author goes on to develop and evaluate the Language Identification followed by Monolingual phone recognition (LID-Mono) and common multilingual phone-set based multilingual phone recognition systems.

**Automatic Speech Recognition on Vibrocervigraphic and Electromyographic Signals**

John Wiley & Sons

Research in the field of automatic speech and speaker recognition has made a number of significant advances in the last two decades, influenced by advances in signal processing, algorithms, architectures, and hardware. These advances include: the adoption of a statistical pattern recognition paradigm; the use of the hidden Markov modeling framework to characterize both the spectral and the temporal variations in the speech signal; the use of a large set of speech utterance examples from a large population of speakers to train the hidden Markov models of some fundamental speech units; the organization of speech and language knowledge sources into a structural finite state network; and the use of dynamic, programming based heuristic search methods to find the best word sequence in the lexical network corresponding to the spoken utterance. Automatic Speech and Speaker Recognition: Advanced Topics groups together in a single volume a number of important topics on speech and speaker recognition, topics which are of fundamental importance, but not yet covered in detail in existing textbooks. Although no explicit partition is given, the book is divided into five parts: Chapters 1-2 are devoted to technology overviews; Chapters 3-12 discuss acoustic modeling of fundamental speech units and lexical modeling of words and pronunciations; Chapters 13-15 address the issues related to flexibility and robustness; Chapter 16-18 concern the theoretical and practical issues of search; Chapters 19-20 give two examples of algorithm and implementational aspects for recognition system realization. Audience: A reference book for speech researchers and graduate students interested in pursuing potential research on the topic. May also be used as a text for advanced courses on the subject.

**Designing Voice User Interfaces** Springer Research is summarized on problems concerned with the interpretation of the acoustic parameters of speech and with the use of linguistic information in processing the discrete code derived from these parameters. The summary discusses work on: the theory of phonology, including development of a phonetic theory based on physiological parameters, and a phonemic theory in natural language and mathematical forms; procedures for converting acoustic parameters to phone types or sets of phone types that can be specified on a phonetic basis; a structural description of the phonology of midwest American English dialect; and lexical procedures for automatic speech recognition. (Author).

*Spoken Language Understanding* Springer

Voice user interfaces (VUIs) are becoming all the rage today. But how do you build one that people can actually converse with? Whether you ’ re designing a mobile app, a toy, or a device such as a home assistant, this practical book guides you through basic VUI design principles, helps you choose the right speech recognition engine, and shows you how to measure your VUI ’ s performance and improve upon it. Author Cathy Pearl also takes product managers, UX designers, and VUI designers into advanced design topics that will help make your VUI not just functional, but great.Understand key VUI design concepts, including command-and-control and conversational systemsDecide if you should use an avatar or other visual representation with your VUIExplore speech recognition technology and its impact on your designTake your VUI above and beyond the basic exchange of informationLearn practical ways to test your VUI application with usersMonitor your app and learn how to quickly improve performanceGet real-world examples of VUIs for home assistants, smartwatches, and car systems