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

Eventually, you will unquestionably discover a additional experience and feat by spending more cash. yet when? complete you acknowledge that you require to get those all needs similar to having significantly cash? Why dont you attempt to acquire something basic in the beginning? Thats something that will guide you to comprehend even more in relation to the globe, experience, some places, next history, amusement, and a lot more?

It is your very own era to be in reviewing habit. along with guides you could enjoy now is **Automatic Speech Recognition On Mobile Devices And Over Communication Networks Advances In Computer Vision And Pattern Recognition** below.

G07H8W - ASHTYN CAYDEN

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.--
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
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."
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.

"This book introduces the readers to the various aspects of visual speech recognitions, including lip segmentation from video sequence, lip feature extraction and modeling, feature fusion and classifier design for visual speech recognition and speaker verification" résumé de l'éditeur.

This book provides a comprehensive overview of the recent advancement in the field of automatic speech recognition with a focus on deep learning models including deep neural networks and many of their variants. This is the first automatic speech recognition book dedicated to the deep learning approach. In addition to the rigorous mathematical treatment of the subject, the book also presents insights and theoretical foundation of a series of highly successful deep learning models.
Automatic Speech Recognition (ASR) is one of the most important applications in the area of cognitive computing. Fast and accurate ASR is emerging as a key application for mobile and wearable devices. These devices, such as smartphones, have incorporated speech recognition as one of the main interfaces for user interaction. This trend towards voice-based user interfaces is likely to continue in the next years which is changing the way of human-machine interaction. Effective speech recognition systems require real-time recognition, which is challenging for mobile devices due to the compute-intensive nature of the problem and the power constraints of such systems and involves a huge effort for CPU architectures to reach it. GPU architectures offer parallelization capabilities which can be exploited to increase the performance of speech recognition systems. However, efficiently utilizing the GPU resources for speech recognition is also challenging, as the software implementations exhibit irregular and unpredictable memory accesses and poor temporal locality. The purpose of this thesis is to study the characteristics of ASR systems running on low-power mobile devices in order to propose different techniques to improve performance and energy consumption. We propose several software-level optimizations driven by the power/performance analysis. Unlike previous proposals that trade accuracy for performance by reducing the number of Gaussians evaluated, we maintain accuracy and improve performance by effectively using the underlying CPU microarchitecture. We use a refactored implementation of the G03 evaluation code to ameliorate the impact of branches. Then, we exploit the vector unit available on most modern CPUs to boost G03 computation, introducing a novel memory layout for storing the means and variances of the Gaussians in order to maximize the effectiveness of vectorization. In addition, we compute the Gaussians for multiple frames in parallel, significantly reducing memory bandwidth usage. Our experimental results show that the proposed optimizations provide 2.68x speedup over the baseline Pocketsphinx decoder on a high-end Intel Skylake CPU, while achieving 61% energy savings. On a modern ARM Cortex-A57 mobile processor our techniques improve performance by 1.85x, while providing 59% energy savings without any loss in the accuracy of the ASR system. Secondly, we propose a register renaming technique that exploits register reuse to reduce the pressure on the register file. Our technique leverages physical register sharing by introducing minor changes in the register map table and the issue queue. We evaluated our renaming technique on top of a modern out-of-order processor. The proposed scheme supports precise exceptions and we show that it results in 9.5% performance improvements for G03 evaluation. Our experimental results show that the proposed register renaming scheme provides 6% speedup on average for the SPEC2006 benchmarks. Alternatively, our renaming scheme achieves the same performance while reducing the number of physical registers by 10.5%. Finally, we propose a hardware accelerator for G03 evaluation that reduces the energy consumption by three orders of magnitude compared to solutions based on CPUs and GPUs. The proposed accelerator implements a lazy evaluation scheme where Gaussians are computed on demand, avoiding 50% of the computations. Furthermore, it employs a novel clustering scheme to reduce the size of the G03 parameters, which results in 8x memory bandwidth savings with a negligible impact on accuracy. Finally, it includes a novel memoization scheme that avoids 74.88% of floating-point operations. The end design provides a 164x speedup and 3532x energy reduction when compared with a highly-tuned implementation running on a modern mobile CPU. Compared to a state-of-the-art mobile GPU, the G03 accelerator achieves 5.89x speedup over a highly optimized CUDA implementation, while reducing energy by 241x.
Automatic speech recognition (ASR) systems are findingincreasing use in everyday life. Many of the commonplaceenvironments where the systems are used are noisy, for exampleusers calling up a voice search system from a busy cafeteria or astreet. This can result in degraded speech recordings and adverselyaffect the performance of speech recognition systems. As theuse of ASR systems increases, knowledge of the state-of-the-art intechniques to deal with such problems becomes critical to systemand application engineers and researchers who work with or on ASRtechnologies. This book presents a comprehensive survey of thestate-of-the-art in techniques used to improve the robustness ofspeech recognition systems to these degrading externalinfluences. Key features: Reviews all the main noise robust ASR approaches, includingsignal separation, voice activity detection, robust featureextraction, model compensation and adaptation, missing datatechniques and recognition of reverberant speech. Acts as a timely exposition of the topic in light of morewidespread use in the future of ASR technology in challengingenvironments. Addresses robustness issues and signal degradation which areboth key requirements for practitioners of ASR. Includes contributions from top ASR researchers from leadingresearch units in the field

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 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.
New material treats such contemporary subjects as automatic speech recognition and speaker verification for banking by computer and privileged (medical, military, diplomatic) information and control access. The book also focuses on speech and audio compression for mobile communication and the

Internet. The importance of subjective quality criteria is stressed. The book also contains introductions to human monaural and binaural hearing, and the basic concepts of signal analysis. Beyond speech processing, this revised and extended new edition of Computer Speech gives an overview of natural language technology and presents the nuts and bolts of state-of-the-art speech dialogue systems.

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.

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.

"Mobile Speech and Advanced Natural Language Solutions" presents the discussion of the most recent advances in intelligent human-computer interaction, including fascinating new study findings on talk-in-interaction, which is the province of conversation analysis, a subfield in sociology/sociolinguistics, a new and emerging area in natural language understanding. Editors Amy Neustein and Judith A. Markowitz have recruited a talented group of contributors to introduce the next generation natural language technologies for practical speech processing applications that serve the consumer's need for well-functioning natural language-driven personal assistants and other mobile devices, while also addressing business' need for better functioning IVR-driven call centers that yield a more satisfying experience for the caller. This anthology is aimed at two distinct audiences: one consisting of speech engineers and system developers; the other comprised of linguists and cognitive scientists. The text builds on the experience and knowledge of each of these audiences by exposing them to the work of the other.

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.

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.

The term Intelligent Environments (IEs) refers to physical spaces in which IT and other pervasive computing technologies are combined and used to achieve specific goals for the user, the environment, or both. The ultimate objective of IEs is to enrich user experience, improve management of the environment in question and increase user awareness. This book presents the proceedings of the following workshops, which formed part of the 12th International Conference on Intelligent Environments (IE16), held in London, UK, in September 2016: the 5th International Workshop on Smart Offices and Other Workplaces (SOOW'16); the 5th International Workshop on the Reliability of Intelligent Environments (WoRIE'16); the 1st International Workshop on Legal Issues in Intelligent Environments (LIIE'2016); the 2nd International Symposium on Future Intelligent Educational Environments and Learning (SOFIEE'16); the 2nd International Workshop on Future Internet and Smart Networks (FI&SN'2016); the International Workshop on Intelligent Environments Supporting Healthcare and Well-being (WISHWell'2016); the International Workshop on Computation Sustainability, Technologies and Applications (CoSTA'2016); the Creative Science 2016 (CS'16) and Cloud-of-Things 2016 (CoT'16); the Workshop on Wireless Body Area Networks for Personal Monitoring in Intelligent Environments (WBAN-PMIE); and the Physical Computing Workshop. The workshops focused on the development of advanced intelligent environments, as well as newly emerging and rapidly evolving topics, emphasizing the multi-disciplinary and transversal aspects of IEs, as well as cutting-edge topics. The book will be of interest to all those whose work involves them in the use of intelligent environments.

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 (ASR) systems are finding increasing use in everyday life. Many of the commonplace environments where the systems are used are noisy, for example users calling up a voice search system from a busy cafeteria or a street. This can result in degraded speech recordings and adversely affect the performance of speech recognition systems. As the use of ASR systems increases, knowledge of the state-of-the-art in techniques to deal with such problems becomes critical to system and application engineers and researchers who work with or on ASR technologies. This book presents a comprehensive survey of the state-of-the-art in techniques used to improve the robustness of speech recognition systems to these degrading external influences. Key features: Reviews all the main noise robust ASR approaches, including signal separation, voice activity detection, robust feature extraction, model compensation and adaptation, missing data techniques and recognition of reverberant speech. Acts as a timely exposition of the topic in light of more widespread use in the future of ASR technology in challenging environments. Addresses robustness issues and signal degradation which are both key requirements for practitioners of ASR. Includes contributions from top ASR researchers from leading research units in the field

This book discusses the contribution of articulatory and excitation source information in discriminating sound units. The authors focus on excitation source component of speech -- and the dynamics of various articulators during speech production -- for enhancement of speech recognition (SR) performance. Speech recognition is analyzed for read, extempore, and conversation modes of speech. Five groups of articulatory features (AFs) are explored for speech recognition, in addition to conventional spectral features. Each chapter provides the motivation for exploring the specific feature for SR task, discusses the methods to extract those features, and finally suggests appropriate models to capture the sound unit specific knowledge from the proposed features. The authors close by discussing various combinations of spectral, articulatory and source features, and the desired models to enhance the performance of SR systems.

Automatic speech recognition systems have to handle various kinds of variabilities sufficiently well in order to achieve high recognition rates in practice. One of the variabilities that has a major impact on the performance is the vocal tract length of the speakers. Normalization of the features and adaptation of the acoustic models are commonly used methods in speech recognition systems. In contrast to that, a third approach follows the idea of extracting features with transforms that are invariant to vocal tract lengths changes. This work presents several approaches for extracting invariant features for automatic speech recognition systems. The robustness of these features under various training-test conditions is evaluated and it is described how the robustness of the features to noise can be increased. Furthermore, it is shown how the spectral effects due to different vocal tract lengths can be estimated with a registration method and how this can be used for speaker normalization.

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.

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.

This book constitutes the refereed proceedings of the Second International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2011, held in Toulouse, France, in August/September 2011. The 30 revised full papers presented were carefully reviewed and selected from numerous submissions. Among the topics addressed are aspects of security, reliability, privacy and anonymity of e-government systems, knowledge processing, service-oriented computing, and case studies of e-government systems in several countries.

This book comprises select papers presented at the International Conference on Mechanical Engineering Design (ICMechD) 2019. The volume focuses on the recent trends in design research and their applications across the mechanical and biomedical domain. The book covers topics like tribology design, mechanism and machine design, wear and surface engineering, vibration and noise engineering, biomechanics and biomedical engineering, industrial thermodynamics, and thermal engineering. Case studies citing practical challenges and their solutions using appropriate techniques and modern engineering tools are also discussed. Given its contents, this book will prove useful to students, researchers as well as practitioners.

Automatic speech recognition (ASR) is a very attractive means for human-machine interaction. The degree of maturity reached by speech recognition technologies during recent years allows the development of applications that use them. In particular, ASR shows an enormous potential in mobile environments, where devices such as mobile phones or PDAs are used, and for Internet Protocol (IP) applications. Speech Recognition Over Digital Channels is the first book of its kind to offer a complete system comprehension, addressing the topics of distributed and network-based speech recognition issues and standards, the concepts of speech processing and transmission, and system architectures and robustness. Describes the different client/server architectures for remote speech recognition systems, by means of which the client transmits speech parameters through a digital channel to a remote recognition server Focuses on robustness against both adverse acoustic environments (in the front-end) and bit errors/packet loss Discusses four ETSI standards for distributed speech recognition; the understanding of the standards and the technologies behind them Provides the necessary background for the comprehension of remote speech recognition technologies This book will appeal to a wide-ranging audience: engineers using speech recognition systems, researchers involved in ASR systems and those interested in processing and transmitting speech such as signal processing and communications communities. It will also be of interest to technical experts requiring an understanding of recognition over mobile and IP networks, and postgraduate students working on robust speech processing.

Two Top Industry Leaders Speak Out Judith Markowitz When Amy asked me to co-author the foreword to her new book on advances in speech recognition, I was honored. Amy's work has always been infused with c- ative intensity, so I knew the book would be as interesting for established speech professionals as for readers new to the speech-processing industry. The fact that I would be writing the foreward with Bill Scholz made the job even more enjoyable. Bill and I have known each other since he was at UNISYS directing projects that had a profound impact on speech-recognition tools and applications. Bill Scholz The opportunity to prepare this foreword with Judith provides me with a rare oppor- nity to collaborate with a seasoned speech professional to identify numerous signi- cant contributions to the field offered by the contributors whom Amy has recruited. Judith and I have had our eyes opened by the ideas and analyses offered by this collection of authors. Speech recognition no longer needs be relegated to the ca- gory of an experimental future technology; it is here today with sufficient capability to address the most challenging of tasks. And the point-click--type approach to GUI control is no longer sufficient, especially in the context of limitations of mode-day hand held devices. Instead, VUI and GUI are being integrated into unified multimodal solutions that are maturing into the fundamental paradigm for comput- human interaction in the future.

The 8th ERCIM Workshop "User Interfaces for All" was held in Vienna, Austria, on 28–29 June 2004, building upon the results of the seven previous workshops held in Heraklion, Crete, Greece, 30–31 October 1995; Prague, Czech Republic, 7–8 Nov- ber 1996; Obernai, France, 3–4 November 1997; Stockholm, Sweden, 19–21 October 1998; Dagstuhl, Germany, 28 November – 1 December 1999; Florence, Italy, 25–26 October 2000; and Paris (Chantilly), France, 24–25 October 2002. The concept of "User Interfaces for All" targets a proactive realization of the "- signforall"principleinthe?eldofhuman-computerinteraction(HCI),andinvolvesthe developmentof user interfaces to interactiveapplicationsand e-services, which provide universalaccess andusabilityto potentiallyall users. In thetraditionofits predecessors, the 8th ERCIM Workshop "User Interfaces for All" aimed to consolidate recent work and to stimulate further discussion on the state of the art in "User Interfaces for All" and its increasing range of applications in the upcoming Information Society. The emphasis of the 2004 event was on "User-Centered Interaction Paradigms for Universal Access in the Information Society. " The requirement for user-centered u- versal access stems from the growing impact of the fusion of the emerging techno- gies and from the different dimensions of diversity that are intrinsic to the Information Society. These dimensions become evident when considering the broad range of user characteristics, the changing nature of human activities, the variety of contexts of use, the increasing availability and diversi?cation of information, knowledge sources and e-services, the proliferation of technological platforms, etc.

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.

The accessibility of mobile terminals, which tend to be lighter and smaller, hampers the development of new services over wireless networks. This trend makes it more difficult, or even frustrates, the interaction of the user with the service. Thus, the development of new user interfaces, providing a ubiquitous, pervasive and multimodal interaction, is a necessary step for the next generation of mobile services. In this scene, automatic speech recognition is a promising way for an easy and natural user access to network services. However, mobile devices are characterized by a restricted computing power, small limited-speed memories and short battery life. In this work, we show how speech recognition based on VoIP technologies allows circumventing these hardware constraints by moving the most complex computational tasks of speech recognition to a remote server. Under this approximation, the user device has to send coded speech or speech parameters through IP networks, which were not designed for real-time communications. For this reason, special emphasis is placed on proposing efficient techniques to avoid the negative impact of network impairments on speech recognition performance.

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.

This book covers language modeling and automatic speech recognition for inflective languages (e.g. Slavic languages), which represent roughly half of the languages spoken in Europe. These languages do not perform as well as English in speech recognition systems and it is therefore harder to develop an application with sufficient quality for the end user. The authors describe the most important language features for the development of a speech recognition system. This is then presented through the analysis of errors in the system and the development of language models and their inclusion in speech recognition systems, which specifically address the errors that are relevant for targeted applications. The error analysis is done with regard to morphological characteristics of the word in the recognized sentences. The book is oriented towards speech recognition with large vocabularies and continuous and even spontaneous speech. Today such applications work with a rather small number of languages compared to the number of spoken languages.

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.

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).

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.