# Robust Automatic Speech Recognition A Bridge To Practical Applications

New Era for Robust Speech Recognition
6th International Conference, SLSP 2018, Mons, Belgium, October 15–16, 2018, Proceedings
Robustness in Automatic Speech Recognition
Computational Auditory Scene Analysis
Proceedings of the International Conference ICCAE, Taipei, Taiwan, November 4-6, 2016
Novel Speech Processing Techniques for Robust Automatic Speech Recognition
Automatic Speech Recognition
Robust Speech
A Study on Noise Robust Acoustic Analysis for Automatic Speech Recognition
Audio Source Separation and Speech Enhancement
Auditory Modeling Inspired Methods of Feature Extraction for Robust Automatic Speech Recognition
Advanced Topics
Auditory Based Modification of MFCC Feature Extraction for Robust Automatic Speech Recognition
Proceedings of the International Conference ICCAE, Taipei, Taiwan, November 4-6, 2016
Noise Reduction in Speech Applications
A Bridge to Practical Applications
A Deep Learning Approach
Robust Speech
Recognition and Understanding
Advancing Noise Robust Automatic Speech Recognition for Command and Control Applications
Informing Multisource Decoding in Robust Automatic Speech Recognition
Robust Adaptation to Non-Native Accents in Automatic Speech Recognition
New Era for Robust Speech Recognition
Theory and Applications
Robust Automatic Speech Recognition and Moduling of Auditory Discrimination with Auditory Experiments Spectro-temporal Features
Recognition and Understanding
Robust Automatic Speech Recognition by Integrating Speech Separation
Robust Speech
Empirical Mode Decomposition for Noise-Robust Automatic Speech Recognition
Robust Speech
Innovative language pedagogy report
Statistical Language and Speech Processing
Robust Automatic Speech Recognition
Robust Speech Recognition of Uncertain or Missing Data
Robust Automatic Speech Recognition with Missing and Unreliable Data
Advancing Noise Robust Automatic Speech Recognition for Command and Control Applications
Studies with Simultaneous Speech
Auditory Front-Ends for Noise-Robust Automatic Speech Recognition
Fundamentals and Applications



## HEAVEN GLASS

*New Era for Robust Speech Recognition* Springer
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
**6th International Conference, SLSP 2018, Mons, Belgium, October 15–16, 2018, Proceedings** IntechOpen
This book on Robust Speech Recognition and Understanding brings together many different aspects of the current research on automatic speech recognition and language understanding. The first four chapters address the task of voice activity detection which is considered an important issue for all speech recognition systems. The next chapters give several extensions to state-of-the-art HMM methods. Furthermore, a number of chapters particularly address the task of robust ASR under noisy conditions. Two chapters on the automatic recognition of a speaker's emotional state highlight the importance of natural speech understanding and interpretation in voice-driven systems. The last chapters of the book address the application of conversational systems on robots, as well as the autonomous acquisition of vocalization skills.
Robustness in Automatic Speech Recognition Springer
This book covers the state-of-the-art in deep neural-network-based methods for noise robustness in distant speech recognition applications. It provides insights and detailed descriptions of some of the new concepts and key technologies in the field, including novel architectures for speech enhancement, microphone arrays, robust features, acoustic model adaptation, training data augmentation, and training criteria. The contributed chapters also include descriptions of real-world applications, benchmark tools and datasets widely used in the field. This book is intended for researchers and practitioners working in the field of speech processing and recognition who are interested in the latest deep learning techniques for noise robustness. It will also be of interest to graduate students in electrical engineering or computer science, who will find it a useful guide to this field of research.
*Computational Auditory Scene Analysis* Springer

Automatic speech recognition (ASR) has been used in many real-world applications such as smart speakers and meeting transcription. It converts speech waveform to text, making it possible for computers to understand and process human speech. When deployed to scenarios with severe noise or multiple speakers, the performance of ASR degrades by large margins. Robust ASR refers to the research field that addresses such performance degradation. Conventionally, the robustness of ASR models to background noise is improved by cascading speech enhancement frontends and ASR backends. This approach introduces distortions to speech signals that can render speech enhancement useless or even harmful for ASR. As for the robustness of ASR models to speech overlaps, traditional frontends cannot use speaker profiles efficiently. In this dissertation, we investigate the integration of ASR backends with speech separation (including speech enhancement and speaker separation) frontends. We start our work by improving the performance of acoustic models in ASR. We propose an utterance-wise recurrent dropout method for a recurrent neural network (RNN) based acoustic model. With utterance-wise context better exploited, the word error rate (WER) reduces substantially. We also propose an iterative speaker adaptation method that can adapt the acoustic model to different speakers using the ASR output from the previous iteration. To obtain a better trade-off between noise reduction and speech distortion for robust monaural (i.e. single-channel) ASR, we train the acoustic model with a large variety of enhanced speech generated by a monaural speech enhancement model. This way, the influence of speech distortion to ASR can be alleviated. We then investigate the use of different types of enhanced features for distortion-independent acoustic modeling. Using distortion-independent acoustic modeling with magnitude features as input, we obtain the state-of-the-art results on the second CHiME speech separation and recognition (CHiME-2) corpus. Multi-channel speech enhancement typically introduces less distortion than monaural speech enhancement. We first substitute the summation operation in beamforming with a learnable complex domain convolutional layer. Operations in complex domain leverage both magnitude and phase information. We then combine this complex domain idea and a two-stage beamforming approach. The first stage extracts spatial features, and the second stage uses both extracted spatial features and the original spectral features as input. This way, the second stage exploits spatial and spectral features explicitly. Using the proposed method, we achieve the state-of-the-art result on the 4th CHiME speech separation and recognition challenge (CHiME-4) corpus. While the enhancement of noisy speech leverages the differences between speech and noise in time-frequency (T-F) patterns, the separation of overlapped speech needs to use speaker-related information. We investigate speaker separation using an inventory of speaker profiles containing speaker identity information. We first select the speaker profiles involved in overlapped speech using an attention-based method. The selected speaker profiles are then used together with the original overlapped speech as input for speaker separation. To alleviate the problem caused by wrong speaker profile selection, we propose to use the output of speaker separation as selected speaker profiles for more iterations of speaker separation. Finally, speech contains sensitive personal data that users may not want to send to cloud-based servers for processing. Next-generation ASR systems should not only be robust to adverse conditions but also lightweight so that they can be deployed on-device. We investigate model compression methods for ASR that do not need model retraining. Our proposed weight sharing based model compression method achieves 9-fold compression with negligible performance degradation.

**Proceedings of the International Conference ICCAE, Taipei, Taiwan, November 4-6, 2016** CRC Press

The 2016 International Conference on Civil, Architecture and Environmental Engineering (ICCAE 2016), November 4-6, 2016, Taipei, Taiwan, is organized by China University of Technology and Taiwan Society of Construction Engineers, aimed to bring together professors, researchers, scholars and industrial pioneers from all over the world. ICCAE 2016 is the premier forum for the presentation and exchange of experience, progress and research results in the field of theoretical and industrial experience. The conference consists of contributions promoting the exchange of ideas between researchers and educators all over the world.

**Novel Speech Processing Techniques for Robust Automatic Speech Recognition** Springer Science & Business Media

Robust Automatic Speech RecognitionA Bridge to Practical ApplicationsAcademic Press

*Automatic Speech Recognition* Academic Press

We present a method for estimating the ASR-driven binary mask using a discriminatively trained sequence model. The hypotheses generated by a first pass through a recognition system are used to estimate the mask. Our proposed method significantly outperforms several methods for estimating the ideal binary mask. We also outline how the system could be used to iteratively improve the estimation.

**Robust Speech** Springer Science & Business Media

The need for automatic speech recognition systems to be robust with respect to changes in their acoustical environment has become more widely appreciated in recent years, as more systems are finding their way into practical applications. Although the issue of environmental robustness has received only a small fraction of the attention devoted to speaker independence, even speech recognition systems that are designed to be speaker independent frequently perform very poorly when they are tested using a different type of microphone or acoustical environment from the one with which they were trained. The use of microphones other than a "close talking" headset also tends to severely degrade speech recognition -performance. Even in relatively quiet office environments, speech is degraded by additive noise from fans, slamming doors, and other conversations, as well as by the effects of unknown linear filtering arising reverberation from surface reflections in a room, or spectral shaping by microphones or the vocal tracts of individual speakers. Speech-recognition systems designed for long-distance telephone lines, or applications deployed in more adverse acoustical environments such as motor vehicles, factory floors, oroutdoors demand far greaterdegrees ofenvironmental robustness. There are several different ways of building acoustical robustness into speech recognition systems. Arrays of microphones can be used to develop a directionally-sensitive system that resists intelference from competing talkers and other noise sources that are spatially separated from the source of the desired speech signal.

**A Study on Noise Robust Acoustic Analysis for Automatic Speech Recognition** Springer

This two-volume work contains the papers presented at the 2016 International Conference on Civil, Architecture and Environmental Engineering (ICCAE 2016) that was held on 4-6 November 2016 in Taipei, Taiwan. The meeting was organized by China University of Technology and Taiwan Society of Construction Engineers and brought together professors, researchers, scholars and industrial pioneers from all over the world. ICCAE 2016 is an important forum for the presentation of new research developments, exchange of ideas and experience and covers the following subject areas: Structural Science & Architecture Engineering, Building Materials & Materials Science, Construction Equipment & Mechanical Science, Environmental Science & Environmental Engineering, Computer Simulation & Computer and Electrical Engineering.

*Audio Source Separation and Speech Enhancement* Springer Science & Business Media

The domain of speech processing has come to the point where researchers and engineers are concerned with how speech technology can be applied to new products, and how this technology will transform our future. One important problem is to improve robustness of speech processing under adverse conditions, which is the subject of this book. Robust speech processing is a relatively new area which became a concern as technology started moving from laboratory to field applications. A method or an algorithm is robust if it can deal with a broad range of applications and adapt to unknown conditions. Robustness in Automatic Speech Recognition addresses all of the fundamental problems and issues in the area. The book is divided into three parts. The first provides the background necessary for understanding the rest of the material. It also emphasizes the problems of speech production and perception in noise along with popular techniques used in speech analysis and automatic speech recognition. Part Two discusses the problems relevant to robustness in automatic speech recognition and speech-based applications. It emphasizes intra- and inter-speaker variability as well as automatic speech recognition of Lombard, noisy and channel distorted speech. Finally, the third part covers recent advances in the field of robust automatic speech recognition. Audience: An invaluable reference. May be used as a text for advanced courses on the subject.

**Auditory Modeling Inspired Methods of Feature Extraction for Robust Automatic Speech Recognition** BoD – Books on Demand

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.

**Advanced Topics** CRC Press

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.

*Auditory Based Modification of MFCC Feature Extraction for Robust Automatic Speech Recognition* Springer

Noise and distortion that degrade the quality of speech signals can come from any number of sources. The technology and techniques for dealing with noise are almost as numerous, but it is only recently, with the development of inexpensive digital signal processing hardware, that the implementation of the technology has become practical. Noise Reduction in Speech Applications provides a comprehensive introduction to modern techniques for removing or reducing background noise from a range of speech-related applications. Self-contained, it starts with a tutorial-style chapter of background material, then focuses on system aspects, digital algorithms, and implementation. The final section explores a variety of applications and demonstrates to potential users of the technology the results possible with the noise reduction techniques presented. The book offers chapters contributed by international experts, a practical, systems approach, and numerous references. For electrical, acoustics, signal processing, communications, and bioengineers, Noise Reduction in Speech Applications is a valuable resource that shows you how to decide whether noise reduction will solve problems in your own systems and how to make the best use of the technologies available.

**Proceedings of the International Conference ICCAE, Taipei, Taiwan, November 4-6, 2016** 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.

*Noise Reduction in Speech Applications* John Wiley & Sons

Automatic speech recognition (ASR) systems still do not perform as well as human listeners under realistic conditions. The unmatched ability of humans to understand speech in most difficult acoustic conditions originates from the superior properties of their auditory system. The aim of this thesis is to improve the recognition performance of ASR systems in difficult acoustic conditions by carefully integrating auditory signal processing strategies. To this end, the physiologically inspired extraction of spectro-temporal modulation patterns was successfully integrated into the front-end of a standard ASR system. Furhter the joint spectro-temporal processing could be separated into independent temporal and spectral processes. To investigate the reason for the remaining "man-maschine-gap" in recognition performance, a range of critical auditory discrimination tasks were performed using ASR systems. The comparison with empirical data showed the the seperate spectro-temporal modulation front-end provides a

suitable auditory model and revealed the importance of across-frequency processing in speech recognition.
Robust Automatic Speech RecognitionA Bridge to Practical Applications
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.
**A Bridge to Practical Applications** IntechOpen
This is a technical assessment paper intended for use by engineers and research scientist working on the development and integration of Automatic Speech Recognition (ASR), it will cover the state of speech and recognition technologies with emphasis on noise robust command and control (C2) application. The reliable elimination of the keyboard and mouse in mounted and un-mounted C2 systems has been a desire of systems developers and requirements writers since the development of PC-based ASR systems in the early 1990's. However, current research and commercial quality ASR applications never had the noise robustness to support a truly tactical C2 application. As ASR achieved limited operational success in noisy environments around the 2002 timeframe, the C2 requirements evolved to include the emerging system of systems approach and multilingual operational environments in support of the Global War On Terrorism (GWOT) in such environment's, the system must understand not just words as commands (ASR), but to understand phrases and sentences (semantic and syntactic) and reply in a conversational manner (speech and natural language generation). If the keyboard and mouse are to be truly eliminated, a system now needs to conduct a natural conversation with an operator and possibly others in the operational environment. This paper will cover the advances, limitations, and reasonable expectations from several levels: Research Scientist and Engineers, Program Executive Office (PEO), Program Manager (PM), and requirements office. I will also discuss the major technical challenges that remain as well as some risk assessment to help decision makers align expectations with reasonable availability dates based on current and future research efforts.
*A Deep Learning Approach* IntechOpen
This is a technical assessment paper intended for use by engineers and research scientist working on the development and integration of Automatic Speech Recognition (ASR). It will cover the state of speech and recognition technologies with emphasis on noise robust command and control (C2) application. The reliable elimination of the keyboard and mouse in mounted and un-mounted C2 systems has been a desire of systems developers and requirements writers since the development of PC-based ASR systems in the early 1990's. However, current research and commercial quality ASR applications never had the noise robustness to support a truly tactical C2 application. As ASR achieved limited operational success in noisy environments around the 2002 timeframe, the C2 requirements evolved to include the emerging system of systems approach and multilingual operational environments in support of the Global War On Terrorism (GWOT) in such environment's, the system must understand not just words as commands (ASR), but to understand phrases and sentences (semantic and syntactic) and reply in a conversational manner (speech and natural language generation). If the keyboard and mouse are to be truly eliminated, a system now needs to conduct a natural conversation with an operator and possibly others in the operational environment. This paper will cover the advances, limitations, and reasonable expectations from several levels: Research Scientist and Engineers, Program Executive Office (PEO), Program Manager (PM), and requirements office. I will also discuss the major technical challenges that remain as well as some risk assessment to help decision makers align expectations with reasonable availability dates based on current and future research efforts.
**Robust Speech** Springer
This book covers the state-of-the-art in deep neural-network-based methods for noise robustness in distant speech recognition applications. It provides insights and detailed descriptions of some of the new concepts and key technologies in the field, including novel architectures for speech enhancement, microphone arrays, robust features, acoustic model adaptation, training data augmentation, and training criteria. The contributed chapters also include descriptions of real-world applications, benchmark tools and datasets widely used in the field. This book is intended for researchers and practitioners working in the field of speech processing and recognition who are interested in the latest deep learning techniques for noise robustness. It will also be of interest to graduate students in electrical engineering or computer science, who will find it a useful guide to this field of research.
*Recognition and Understanding* Research-publishing.net
This book on Robust Speech Recognition and Understanding brings together many different aspects of the current research on automatic speech recognition and language understanding. The first four chapters address the task of voice activity detection which is considered an important issue for all speech recognition systems. The next chapters give several extensions to state-of-the-art HMM methods. Furthermore, a number of chapters particularly address the task of robust ASR under noisy conditions. Two chapters on the automatic recognition of a speaker's emotional state highlight the importance of natural speech understanding and interpretation in voice-driven systems. The last chapters of the book address the application of conversational systems on robots, as well as the autonomous acquisition of vocalization skills.

Related with Robust Automatic Speech Recognition A Bridge To Practical Applications:
• Historia De La Familia Greco : click here