Implementing Speech Recognition Algorithms On The

This work assimilates an introduction to the Human speech production system, Fundamentals of speech recognition, the methods for data analysis as well as the theoretical background of past research, Soft computing, implementation as a form of Methodology, Results and Conclusion & future work . Soft Computing includes a family consisting of many members, namely Genetic Algorithms (GAs), Neural Networks and others.With this text, you will find a balance between theory and implementation that allows you to build your understanding of the basic concepts, Methodology and a results obtained using various neural networks with Recognition percentages of isolated words in two different languages and Confusion Matrix for radial basis similarly for probabilistic. This work also shows the Comparative analysis of results from the Recognition using GA and without GA. The use of GAs for feature optimizers is a novel concept in Automatic Speech Recognition hence the system becomes more robust to noise. Emphasis is laid more on a soft computing rather than on a hard one.
Contents:A Connectionist Approach to Speech Recognition (Y Bengio)Signature Verification Using a “Siamese” Time Delay Neural Network (J Bromley et al.)Boosting Performance in Neural Networks (H Drucker et al.)An Integrated Architecture for Recognition of Totally Unconstrained Handwritten Numerals (A Gupta et al.)Time-Warping Network: A Neural Approach to Hidden Markov Model Based Speech Recognition (E Levin et al.)Computing Optical Flow with a Recurrent Neural Network (H Li & J Wang)Integrated Segmentation and Recognition through Exhaustive Scans or Learned Saccadic Jumps (G L Martin et al.)Experimental Comparison of the Effect of Order in Recurrent Neural Networks (C B Miller & C L Giles)Adaptive Classification by Neural Net Based Prototype Populations (K Peleg & U Ben-Hanan)A Neural System for the Recognition of Partially Occluded Objects in Cluttered Scenes: A Pilot Study (L Wiskott & C von der Malsburg)and other papers Readership: Computer scientists and engineers.
This textbook explains Deep Learning Architecture, with applications to various NLP Tasks, including Document Classification, Machine Translation, Language Modeling, and Speech Recognition. With the widespread adoption of deep learning, natural language processing (NLP),and speech applications in many areas (including Finance, Healthcare, and Government) there is a growing need for one comprehensive resource that maps deep learning techniques to NLP and speech and provides insights into using the tools and libraries for real-world applications. Deep Learning for NLP and Speech Recognition explains recent deep learning methods applicable to NLP and speech, provides state-of-the-art approaches, and offers real-world case studies with code to provide hands-on experience. Many books focus on deep learning theory or deep learning for NLP-specific tasks while others are cookbooks for tools and libraries, but the constant flux of new algorithms, tools, frameworks, and libraries in a rapidly evolving landscape means that there are few available texts that offer the material in this book. The book is organized into three parts, aligning to different groups of readers and their expertise. The three parts are: Machine Learning, NLP, and Speech Introduction The first part has three chapters that introduce readers to the fields of NLP, speech recognition, deep learning and machine learning with basic theory and hands-on case studies using Python-based tools and libraries. Deep Learning Basics The five chapters in the second part introduce deep learning and various topics that are crucial for speech and text processing, including word embeddings, convolutional neural networks, recurrent neural networks and speech recognition basics. Theory, practical tips, state-of-the-art methods, experimentations and analysis in using the methods discussed in theory on real-world tasks. Advanced Deep Learning Techniques for Text and Speech The third part has five chapters that discuss the latest and cutting-edge research in the areas of deep learning that intersect with NLP and speech. Topics including attention mechanisms, memory augmented networks, transfer learning, multi-task learning, domain adaptation, reinforcement learning, and end-to-end deep learning for speech recognition are covered using case studies.
Implementation of I-vector Algorithm in Speech Emotion Recognition by Using Two Different Classifiers
Speech Recognition Using Vector Quantization
Gaussian Mixture Model and Support Vector Machine
A Bridge to Practical Applications
Computer Models of Speech Using Fuzzy Algorithms

*Automatic Speech Recognition (ASR) is the enabling technology for hands-free dictation and voice-triggered computer menus. It is becoming increasingly prevalent in environments such as private telephone exchanges and real-time information services. Speech Recognition introduces the principles of ASR systems, including the theory and implementation issues behind multi-speaker continuous speech recognition. Focusing on the algorithms employed in commercial and laboratory systems, the treatment enables the reader to devise practical solutions for ASR system problems. It addresses in detail C++ programming techniques used to develop ASR applications, thus offering skills that will prove useful in any large C++ based software project. Possible extensions of the well-established ASR technology are highlighted, based on "Hidden Markov Models" applied to fields such as modelling and prediction of econometric series. Features include: * Accompanying website containing all C++ source code of a complete laboratory multi-speaker continuous-speech ASR system (e.g. Initialisation, Training, Recognition, Evaluation, etc.) www.wiley.com/go/becchetti_speech * Detailed theoretical, mathematical and technical explanations of ASR * A practical account of the functioning of ASR A crucial source of information for researchers, developers and project managers involved with ASR systems, Speech Recognition is also structured for use by students of digital signal processing, speech recognition and C++ programming techniques.*
Emotions are essential for our existence, as they exert great influence on the mental health of people. Speech is the most powerful mode to communicate. It controls our intentions and emotions. Over the past years many researchers worked hard to recognize emotion from speech samples. Many systems have been proposed to make the Speech Emotion Recognition (SER) process more correct and accurate. This thesis research discusses the design of speech emotion recognition system implementing a comparatively new method, i-vector model. I-vector model has found much success in the areas of speaker identification, speech recognition, and language identification. But it has not been much explored in recognition of emotion. In this research, i-vector model was implemented in processing extracted features for speech representation. Two different classification schemes were designed using two different classifiers - Gaussian Mixture Model (GMM) and Support Vector Machine (SVM), along with i-vector algorithm. Performance of these two systems was evaluated using the same emotional speech database to identify four emotional speech signals: Angry, Happy, Sad and Neutral. Results were analyzed, and more than 75% of accuracy was obtained by both systems, which proved that our proposed i-vector algorithm can identify speech emotions with less error and with more accuracy.
The paper describes techniques and methodology which are useful in achieving close to real-time recognition of speech by a computer. To analyze connected speech utterances, any speech recognition system must perform the following processes: preprocessing, segmentation, segment classification, recognition of words, recognition of sentences. The paper presents implemented solutions to each of these problems which achieved accurate recognition in all the trial cases. (Author).
Dynamic Speech Models
Deep Learning with Applications Using Python
A Parallel Implementation of the A-Viterbi Algorithm for Speech Recognition*
Design, Implementation, and Samples Codes
From Natural to Artificial Intelligence

Speech recognition algorithms were analyzed using normal and G-stressed speech as an input. Speech samples were recorded in centrifuge tests at the Air Force Medical Research Lab, Wright-Patterson AFB, Ohio. All speech was recorded using the MBU-12/P face mask. The algorithms studied are phoneme-based feature extractors which feed a recognition algorithm based on fuzzy set theory. Three feature extraction algorithm options were analyzed. One option used a phoneme length of 40 ms and the other options used a length of 8 ms. The recognition results for all three options using normal speech are above 90%, but the 40ms phoneme length give higher raw scores. For G-stressed speech the 40 ms phoneme length scored greater than 90% while the 8ms phoneme length options scored less than 60%. (Author).
"While deep learning algorithms have made significant progress in automatic speech recognition and natural language processing, they require a significant amount of labelled training data to perform effectively. As such, these applications have not been extended to languages that have only limited amount of data available, such as extinct or endangered languages. Another problem caused by the rise of deep learning is that individuals with malicious intents have been able to leverage these algorithms to create fake contents that can pose serious harm to security and public safety. In this work, we explore the solutions to both of these problems. First, we investigate different data augmentation methods and acoustic architecture designs to improve automatic speech recognition performance on low-resource languages. Data augmentation for audio often involves changing the characteristic of the audio without modifying the ground truth. For example, different background noise can be added to an utterance while maintaining the content of the speech. We also explored how different acoustic model paradigms and complexity affect performance on low-resource languages. These methods are evaluated on Seneca, an endangered language spoken by a Native American tribe, and Iban, a low-resource language spoken in Malaysia and Brunei. Secondly, we explore methods to determine speaker identification and audio spoofing detection. A spoofing attack involves using either a text-to-speech voice conversion application to generate audio that mimic the identity of a target speaker. These methods are evaluated on the ASVSpoof 2019 Logical Access dataset containing audio generated using various methods of voice conversion and text-to-speech synthesis."--Abstract.
Speech Processing has rapidly emerged as one of the most widespread and well-understood application areas in the broader discipline of Digital Signal Processing. Besides the telecommunications applications that have hitherto been the largest users of speech processing algorithms, several non-traditional embedded processor applications are enhancing their functionality and user interfaces by utilizing various aspects of speech processing. "Speech Processing in Embedded Systems" describes several areas of speech processing, and the various algorithms and industry standards that address each of these areas. The topics covered include different types of Speech Compression, Echo Cancellation, Noise Suppression, Speech Recognition and Speech Synthesis. In addition this book explores various issues and considerations related to efficient implementation of these algorithms on real-time embedded systems, including the role played by processor CPU and peripheral functionality.
Implementation of a Speech Recognition Algorithm Utilizing the TMS320 Digital Signal Processor
Theory and C++ Implementation
A Real-time, Vectorized, Large Vocabulary Speech Recognition Algorithm Using Parallel Processing
Robust Automatic Speech Recognition
A Discriminative Training Algorithm and Its Implementation to Automatic Speech Recognition

This book introduces the theory, algorithms, and implementation techniques for efficient decoding in speech recognition mainly focusing on the Weighted Finite-State Transducer (WFST) approach. The decoding process for speech recognition is viewed as a search problem whose goal is to find a sequence of words that best matches an input speech signal. Since this process becomes computationally more expensive as the system vocabulary size increases, research has long been devoted to reducing the computational cost. Recently, the WFST approach has become an important state-of-the-art speech recognition technology, because it offers improved decoding speed with fewer recognition errors compared with conventional methods. However, it is not easy to understand all the algorithms used in this framework, and they are still in a black box for many people. In this book, we review the WFST approach and aim to provide comprehensive interpretations of WFST operations and decoding algorithms to help anyone who wants to understand, develop, and study WFST-based speech recognizers. We also mention recent advances in this framework and its applications to spoken language processing. Table of Contents: Introduction / Brief Overview of Speech Recognition / Introduction to Weighted Finite-State Transducers / Speech Recognition by Weighted Finite-State Transducers / Dynamic Decoders with On-the-fly WFST Operations / Summary and Perspective
It is with great pleasure that I present this third volume of the series "Advanced Applications in Pattern Recognition." It represents the summary of many man- (and woman-) years of effort in the field of speech recognition by tne author's former team at the University of Turin. It combines the best results in fuzzy-set theory and artificial intelligence to point the way to definitive solutions to the speech-recognition problem. It is my hope that it will become a classic work in this field. I take this opportunity to extend my thanks and appreciation to Sy Marchand, Plenum's Senior Editor responsible for overseeing this series, and to Susan Lee and Jo Winton, who had the monumental task of preparing the camera-ready master sheets for publication. Morton Nadler General Editor vii PREFACE Si parva licet componere magnis Virgil, Georgics, 4,176 (37-30 B.C.) The work reported in this book results from years of research oriented toward the goal of making an experimental model capable of understanding spoken sentences of a natural language. This is, of course, a modest attempt compared to the complexity of the functions performed by the human brain. A method is introduced for conce1v1ng modules performing perceptual tasks and for combining them in a speech understanding system.
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
Speech Recognition Using Rector Quantization, Markor Models and the Viterbi Algorithm
Fundamentals of Speech Recognition
Deep Learning for NLP and Speech Recognition
Advanced Architectures for Digital Signal Processors
Speech Recognition Algorithms Based on Weighted Finite-State Transducers

"This thesis comments on the results thus obtained, and attempts to explain the behaviour of the different parallel implementations." --
Abstract: "The purpose of this project is to implement a series of compensation algorithms that would enable the NeXT workstations to make use of the SPHINX recognition system using the microphone that is mounted in the bezel of the NeXT Station, and to explore the reasons for the degradation of the recognition performance when the microphone mounted in the bezel is used. Currently the performance of SPHINX is inadequate unless an external headset-mounted microphone is used. Sound was acquired using the surface microphone and CODEC chip, and comparisons were made to performance using higher-bandwidth, higher-quality speech input through the DSP port of the machine. In this dissertation we present our findings concerning the performance of the SPHINX automatic speech recognition system when applied to the microphone embedded in the bezel of the NeXT machine. We observed an increase of the relative error rate of 423% when switching from speech recorded in a controlled acoustical environment to the one provided by the NeXT microphone, leaving large improvement opportunities. Several compensation algorithms that model the degradation as the addition of linear filtering and additive noise were applied, including SDCN, FCDCN, and CDCN. The recognition improvement obtained using these algorithms compensated for half the possible upgrade, one that would compare with a system trained and tested on the same microphone. The main reasons for these shortcoming [sic] were the inadequacy to produce stereo databases from simultaneous recording along with heavy environmental distortions."
Explore deep learning applications, such as computer vision, speech recognition, and chatbots, using frameworks such as TensorFlow and Keras. This book helps you to ramp up your practical know-how in a short period of time and focuses you on the domain, models, and algorithms required for deep learning applications. Deep Learning with Applications Using Python covers topics such as chatbots, natural language processing, and face and object recognition. The goal is to equip you with the concepts, techniques, and algorithm implementations needed to create programs capable of performing deep learning. This book covers convolutional neural networks, recurrent neural networks, and multilayer perceptrons. It also discusses popular APIs such as IBM Watson, Microsoft Azure, and scikit-learn. What You Will Learn Work with various deep learning frameworks such as TensorFlow, Keras, and scikit-learn. Use face recognition and face detection capabilities Create speech-to-text and text-to-speech functionality Engage with chatbots using deep learning Who This Book Is For Data scientists and developers who want to adapt and build deep learning applications.
Progressive-Search Algorithms for Large-Vocabulary Speech Recognition
Sub-word based Tigrinya speech recognizer. An experiment using hidden Markov model
Speech Recognition
Design of a Speech Recognition System Using Template-based Dynamic-time-warp Algorithm
Soft Computing Implementation of Automatic Speech Recognition

The authors describe a technique they call "Progressive Search," which is useful for developing and implementing speech recognition systems with high computational requirements. The scheme iteratively uses more and more complex recognition schemes, where each iteration constrains the search space of the next. An algorithm, the "Forward-Backward Word-Life Algorithm," is described. It can generate a word lattice in a progressive search that would be used as a language model embedded in a succeeding recognition pass to reduce computation requirements. They show that speed-ups of more than an order of magnitude are achievable with only minor costs in accuracy.
The goal of SRI's consistency modeling project is to improve the raw acoustic modeling component of SRI's DECIPHER speech recognition system and develop consistency modeling technology. Consistency modeling aims to reduce the number of improper independence assumptions used in traditional speech recognition algorithms so that the resulting speech recognition hypotheses are more self-consistent and, therefore, more accurate. At the initial stages of this effort, SRI focused on developing the appropriate base technologies for consistency modeling. We first developed the Progressive Search technology that allowed us to perform large-vocabulary continuous speech recognition (LVCSR) experiments. Since its conception and development at SRI, this technique has been adopted by most laboratories, including other ARPA contracting sites, doing research on LVSR. Another goal of the consistency modeling project is to attack difficult modeling problems, when there is a mismatch between the training and testing phases. Such mismatches may include outlier speakers, different microphones and additive noise. We were able to either develop new, or transfer and evaluate existing, technologies that adapted our baseline genonic HMM recognizer to such difficult conditions. (AN).
This report is concerned with the development and application of improved techniques for digital signal processing, based on use of residue number system (RNS) to implement the processing functions associated with isolated-word speech recognition. Specifically, the use of RNS in combination and systolic architectures for implementation of speech recognition algorithms was explored. The implementation of time-warping speech algorithm in RNS is described. Keyword include: Signal processing, Speech recognition, systolic architecture, Residue number system (RNS), Autoregressive spectral estimation, and Dynamic time-warping.
Speech Recognition Using Neural Nets and Dynamic Time Warping
Using Genetic Algorithm to Improve the Performance of Speech Recognition Based on Artificial Neural Network
Experimental Evaluation of Algorithms for Connected Speech Recognition Using Hidden Markov Models
Feed-forward Neural Networks for Isolated Speech Recognition Using Four Different Backpropagation Training Algorithms
Theory and Practice

In this book, we introduce the background and mainstream methods of probabilistic modeling and discriminative parameter optimization for speech recognition. The specific models treated in depth include the widely used exponential-family distributions and the hidden Markov model. A detailed study is presented on unifying the common objective functions for discriminative learning in speech recognition, namely maximum mutual information (MMI), minimum classification error, and minimum phone/word error. The unification is presented, with rigorous mathematical analysis, in a common rational-function form. This common form enables the use of the growth transformation (or extended Baum–Welch) optimization framework in discriminative learning of model parameters. In addition to all the necessary introduction of the background and tutorial material on the subject, we also included technical details on the derivation of the parameter optimization formulas for exponential-family distributions, discrete hidden Markov models (HMMs), and continuous-density HMMs in discriminative learning. Selected experimental results obtained by the authors in firsthand are presented to show that discriminative learning can lead to superior speech recognition performance over conventional parameter learning. Details on major algorithmic implementation issues with practical significance are provided to enable the practitioners to directly reproduce the theory in the earlier part of the book into engineering practice. Table of Contents: Introduction and Background / Statistical Speech Recognition: A Tutorial / Discriminative Learning: A Unified Objective Function / Discriminative Learning Algorithm for Exponential-Family Distributions / Discriminative Learning Algorithm for Hidden Markov Model / Practical Implementation of Discriminative Learning / Selected Experimental Results / Epilogue / Major Symbols Used in the Book and Their Descriptions / Mathematical Notation / Bibliography
Research in automatic speech recognition has been done for almost four decades. This project aims to develop automated English digits speech recognition system using Matlab. The system is able to recognize the spoken utterances by translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique, which then estimates the observation likelihood by using the Forward algorithm. The Hidden Markov Model (HMM) parameters are estimated by applying the Baum-Welch algorithm on previously trained samples. The most likely sequence is then decoded using Viterbi algorithm, thus producing the recognized word. This project focuses on all English digits from (Zero through Nine), which is based on isolated words structure. Two modules were developed, namely the isolated words speech recognition and the continuous speech recognition. Both modules were tested in both clean and noisy environments and showed relatively sucessful recognition rates. The samples of Matlab codes were provided in the Appendix.
Abstract: "This thesis examines how artificial neural networks can benefit a large vocabulary, speaker independent, continuous speech recognition system. Currently, most speech recognition systems are based on hidden Markov models (HMMs), a statistical framework that supports both acoustic and temporal modeling. Despite their state-of-the-art performance, HMMs make a number of suboptimal modeling assumptions that limit their potential effectiveness. Neural networks avoid many of these assumptions, while they can also learn complex functions, generalize effectively, tolerate noise, and support parallelism. While neural networks can readily be applied to acoustic modeling, it is not yet clear how they can be used for temporal modeling. Therefore, we explore a class of systems called NN-HMM hybrids, in which neural networks perform acoustic modeling, and HMMs perform temporal modeling. We argue that a NN-HMM hybrid has several theoretical advantages over a pure HMM system, including better acoustic modeling accuracy, better context sensitivity, more natural discrimination, and a more economical use of parameters. These advantages are confirmed experimentally by a NN-HMM hybrid that we developed, based on context-independent phoneme models, that achieved 90.5% word accuracy on the Resource Management database, in contrast to only 86.0% accuracy achieved by a pure HMM under similar conditions. In the course of developing this system, we explored two different ways to use neural networks for acoustic modeling: prediction and classification. We found that predictive networks yield poor results because of a lack of discrimination, but classification networks gave excellent results. We verified that, in accordance with theory, the output activations of a classification network form highly accurate estimates of the posterior probabilities P(class/input), and we showed how these can easily be converted to likelihoods P(input/class) for standard HMM recognition algorithms. Finally, this thesis reports how we optimized the accuracy of our system with many natural techniques, such as expanding the input window size, normalizing the inputs, increasing the number of hidden units, converting the network's output activations to log likelihoods, optimizing the learning rate schedule by automatic search, backpropagating error from word level outputs, and using gender dependent networks."
Discriminative Learning for Speech Recognition
Robust Speech Recognition Using a Noise Rejection Approach
Voice Recognition System in Noisy Environment
A Discriminative Adaptation Algorithm and Its Implementation to Automatic Speech Recognition
Markov Models and the Viterbi Algorithm

Speech dynamics refer to the temporal characteristics in all stages of the human speech communication process. This speech “chain” starts with the formation of a linguistic message in a speaker's brain and ends with the arrival of the message in a listener's brain. Given the intricacy of the dynamic speech process and its fundamental importance in human communication, this monograph is intended to provide a comprehensive material on mathematical models of speech dynamics and to address the following issues: How do we make sense of the complex speech process in terms of its functional role of speech communication? How do we quantify the special role of speech timing? How do the dynamics relate to the variability of speech that has often been said to seriously hamper automatic speech recognition? How do we put the dynamic process of speech into a quantitative form to enable detailed analyses? And finally, how can we incorporate the knowledge of speech dynamics into computerized speech analysis and recognition algorithms? The answers to all these questions require building and applying computational models for the dynamic speech process. What are the compelling reasons for carrying out dynamic speech modeling? We provide the answer in two related aspects. First, scientific inquiry into the human speech code has been relentlessly pursued for several decades. As an essential carrier of human intelligence and knowledge, speech is the most natural form of human communication. Embedded in the speech code are linguistic (as well as para-linguistic) messages, which are conveyed through four levels of the speech chain. Underlying the robust encoding and transmission of the linguistic messages are the speech dynamics at all the four levels. Mathematical modeling of speech dynamics provides an effective tool in the scientific methods of studying the speech chain. Such scientific studies help understand why humans speak as they do and how humans exploit redundancy and variability by way of multitiered dynamic processes to enhance the efficiency and effectiveness of human speech communication. Second, advancement of human language technology, especially that in automatic recognition of natural-style human speech is also expected to benefit from comprehensive computational modeling of speech dynamics. The limitations of current speech recognition technology are serious and are well known. A commonly acknowledged and frequently discussed weakness of the statistical model underlying current speech recognition technology is the lack of adequate dynamic modeling schemes to provide correlation structure across the temporal speech observation sequence. Unfortunately, due to a variety of reasons, the majority of current research activities in this area favor only incremental modifications and improvements to the existing HMM-based state-of-the-art. For example, while the dynamic and correlation modeling is known to be an important topic, most of the systems nevertheless employ only an ultra-weak form of speech dynamics; e.g., differential or delta parameters. Strong-form dynamic speech modeling, which is the focus of this monograph, may serve as an ultimate solution to this problem. After the introduction chapter, the main body of this monograph consists of four chapters. They cover various aspects of theory, algorithms, and applications of dynamic speech models, and provide a comprehensive survey of the research work in this area spanning over past 20~years. This monograph is intended as advanced materials of speech and signal processing for graudate-level teaching, for professionals and engineering practioners, as well as for seasoned researchers and engineers specialized in speech processing
Human life could be much more comfortable, if global machinery works on voice commands. There are lot of products available in the market those provide us automated voice controlled working experience. Voice-identification is a kind of technology in which, machine responds to human voice. The goal of this project is to design and implement voice controlled embedded system in noisy environment. The project consists software and hardware design. The software part is to eliminate noise from the original signal and develop the speech recognition algorithm (DTW). The hardware is used for speech recognition. In this project, DTW algorithm is developed to study and to research the implementation of the speech recognition for single-word. In addition, in this project I implemented nearest neighbor algorithm followed by DTW, which helps to match the speech with different people's accents. For really good result I used wiener filter to increase signal to noise ratio of applied signal. Software portion is fully developed in MATLAB environment. At the other end, hardware part was concentrated on comparing and processing of applied speech with pre-stored speech signal using HM2007. Also after the processing, output of the hardware controls the system according to the speech.
The purpose of this study is to demonstrate the feasibility of using Kohonen neural nets in speech recognition. This is done by combining a first level Kohonen net with a work recognition algorithm which is either dynamic time warping (DTW) or a second Kohonen net. A digitized utterance is sliced and processed to obtain a sequence of 15 component vectors. Each component corresponds to the energy in a selected frequency range. An utterance of the digits zero through nine is used to train the first Kohonen net. After training, an utterance input to the net produces a trajectory through the net. Each point on the trajectory corresponds to a node and a particular sound. These trajectories are input to a work recognition algorithm. The first of these, DTW, compares unknown utterances to template utterances. It is a computationally intense, mathematical algorithm, and it was used primarily to test the preprocessing and neural net training procedures. The second algorithm is a second Kohonen neural net. Digits are assigned to each node so that when an unknown trajectory is input to the second net, the node that lights up identifies the utterance. Using DTW, 99% isolated and 93% connected speech recognition rates are achieved. With the second Kohonen net, isolated speech is recognized at up to 96%, depending upon the net format. Recommendations for future effort include increasing the vocabulary, using multiple feature sets and nets to attempt speaker independent speech recognition, and substituting a backward propagation multi-layer perceptron net for word recognition. Theses. (RH).
Algorithms and Applications
Speech Processing in Embedded Systems
Advances in Pattern Recognition Systems Using Neural Network Technologies
Statistical Methods for Speech Recognition
High-Performance Speech Recognition Using Consistency Modeling

Provides a theoretically sound, technically accurate, and complete description of the basic knowledge and ideas that constitute a modern system for speech recognition by machine. Covers production, perception, and acoustic-phonetic characterization of the speech signal; signal processing and analysis methods for speech recognition; pattern comparison techniques; speech recognition system design and implementation; theory and implementation of hidden Markov models; speech recognition based on connected word models; large vocabulary continuous speech recognition; and task- oriented application of automatic speech recognition. For practicing engineers, scientists, linguists, and programmers interested in speech recognition.
Current Automatic Speech Recognition devices attempt to solve the connected word recognition problem by assuming that an unknown phrase is the output of a sequence of statistical word-models. Typically, these models are constructed using examples of words spoken in isolation; however, the acoustic patterns corresponding to words as they occur in fluent speech are quite different from those representing the same words spoken in isolation, and so the use in speech recognizers of models based on isolated utterances severely limits the performance of such devices. A method of extracting training utterances from fluent speech and constructing Hidden Markov Models (HMMs) from these templates, known as Embedded Training, is investigated here, in conjunction with a two-level algorithm for connected word recognition. The effects on recognition performance of various HMM training procedures are discussed, and experimental results are presented.
Master's Thesis from the year 2013 in the subject Computer Science - Miscellaneous, grade: Very good, , course: Masters of Science in Computer Science, language: English, abstract: Speech recognition, a process of changing speech to text, has been one of a research area for the last many decades. Even though there are several techniques of modeling a speech recognizer, yet it is still challenging to find one that overcomes all the limitations. So this thesis examines the possibility of developing Tigrinya language speech recognizer by finding out which sub-word unit is most appropriate in developing efficient large vocabulary, speaker independent, and continuous Tigrinya speech recognition system using hidden Markov models (HMM). The recognizer was developed using Hidden Markov Model, and the Hidden Markov Modeling Toolkit was used to implement it. In the course of developing this system, the speech data is recorded at a sampling rate of 16 KHz and the recorded speech is converted into Mel Frequency Cepstral Coefficient (MFCC) vectors for further analysis and processing. In this research work, 1000 selected utterances were uttered by 26 selected peoples from different age group and sex constituting of 4643 unique words. Accordingly, the database is set up into two ways the first database comprised of 1000 utterances that are used for training and out of which 100 sentences were taken for testing and evaluation whereas the second database consists of 900 utterances for training and 100 utterances for test and evaluation which is different from the training set. Furthermore, the data is preprocessed in line with the requirements of the HTK toolkit and both the text and speech corpuses were prepared in consultation with the domain experts.
Speech Recognition in Noise Using Weighted Matching Algorithms
Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras
Aspects of Speech Recognition by Computer
Computer Recognition of Phonemes in the Presence of Cockpit Induced Stress and Noise
Speech Recognition Using the NeXT Microphone in the SPHINX System

In this chapter, we have seen that the recognition rate through the SDM in BPNN is up to 91% under the MFCC feature. This recognition rate is not the optimum because that the SDM can always get local optimum. To solve this problem, GA was adopted and following SDM to improve MSE. By this two stage (SDM then GA) training scheme, the recognition rate can be increasing up to 95%. However, under the condition of adopting only MFCC parameters, speech recognition rate still has room for improvement. For the future, other.
This book reflects decades of important research on the mathematical foundations of speech recognition. It focuses on underlying statistical techniques such as hidden Markov models, decision trees, the expectation-maximization algorithm, information theoretic goodness criteria, maximum entropy probability estimation, parameter and data clustering, and smoothing of probability distributions. The author's goal is to present these principles clearly in the simplest setting, to show the advantages of self-organization from real data, and to enable the reader to apply the techniques.
Speech Recognition Using Neural Networks
Speech Recognition System Using MATLAB
Deepfake Detection and Low-resource Language Speech Recognition Using Deep Learning