

Facultatea: Electronică, Telecomunicații și Tehnologia informației

Titlul tezei: Recunoașterea Limbajului Vorbit în Rețele de Telecomunicații Mobile

Autor: Andi Buzo

Conducător de doctorat: Corneliu Burileanu

Abstract

Această teză de doctorat are două obiective principale. Primul obiectiv este dezvoltarea unui sistem de recunoaștere a limbajului vorbit (SRLV) în limba română. Al doilea obiectiv este analiza comportării sistemului în condițiile în care el, sau o parte din el, este implementat pe un dispozitiv portabil, iar semnalul vocal este trimis printr-o rețea de telecomunicații. Teza începe cu o descriere detaliată a stadiului de dezvoltare în domeniul recunoașterii vorbirii, atât pe plan internațional cât și pe plan național. În primele capitole, sunt descriși algoritmi de extragere ai parametrilor vocali, de antrenare și de decodare bazați pe modele markov ascunse. Sunt descriși, de asemenea, algoritmi care oferă o robustețe ridicată a SRLV în condiții de zgomot și pierderi de pachete care se întâlnesc des în rețele de telecomunicații. În această lucrare sunt prezentate metode pentru achiziționarea bazelor de date în limba română, strategia de antrenare și infrastructura SRLV construită cu scopul de a ușura rularea utilitatelor și de a oferi flexibilitate la variația configurației. Un număr mare de experimente sunt făcute cu scopul de a găsi configurația optimă a SRLV în limba română și pentru validarea unor metode de optimizare atât pentru acuratețea recunoașterii cât și pentru timpul de rulare. Sunt propuse niște metode de reducere a timpului de calcul cum ar fi reducerea fascicolului de căutare și micșorarea adaptivă a dicționarului fonetic. În această lucrare, este formulată o teorie bazată pe incertitudine și sunt prezentate rezultatele experimentale pentru validarea teoriei și aplicarea acesteia în cazul condițiilor adverse întâlnite în rețelele de telecomunicații (zgomot și pierderi de pachete). Se definește o mărime nouă numită incertitudinea modelului acustic care se folosește pentru o re-evaluare a modelului acustic. S-au construit două metode pe baza acestei mărimi: una cu scopul de a scoate parametri de voce irelevanți, reducând astfel numărul de calcule necesare, iar cea de-a doua cu scopul de a ameliora rata de recunoaștere. Metodele sunt aplicate atât la nivel de model cât și la nivel de stare. O evaluare a metodelor este făcută și cu baza de date TIMIT cu scopul de a valida generalitatea lor.

Faculty: Electronics, Telecommunications and Information Technology

Ph.D. thesis title: Automatic Speech Recognition over Mobile Communication Networks

Author: Andi Buzo

Supervisor: Corneliu Burileanu

Abstract

This Ph.D. thesis has two main objectives. The first objective is the development of an Automatic Speech Recognition System (ASRS) for the Romanian language. The second objective is the system analysis when the system or part of it is implemented on a mobile device and the speech signal is sent over communication networks. This study starts with a detailed description of both international and national state-of-the-art in the domain of automatic speech recognition. In the first chapters, there are described the algorithms for speech feature extraction, hidden Markov model based training and decoding. The algorithms for ASRS robustness against noise and packet loss (which occur frequently over communication channels) are also described. In this study, there are presented several methods for the Romanian language database acquisition, the training strategy and the ASRS infrastructure built with the purpose of easing the tools/scripts run and increasing the flexibility of system configuration. A large number of experiments are run with the purpose of finding the optimal ASRS configuration for Romanian language and validation of some optimization methods from both recognition accuracy and running time point of view. Several methods are proposed for the decrease of system complexity by reducing the search beam size or by resizing adaptively the phonetic dictionary. In this study, an uncertainty based theory is built and experimental results are given for the validation of the theory and its application in impairment conditions (noise and packet loss) met in communication networks. A new measure is defined: the acoustic model uncertainty which is used for rescoring in the decoding process. Two methods are built based on this new measure: the first one reduces the complexity of the system by removing the irrelevant speech features, while the second one increases the recognition accuracy by weighting the speech features according to their relevance. The methods are applied at both model and state level. The methods are also evaluated with the standard TIMIT database to prove their generality.