

République Algérienne Démocratique et populaire
Université M'hamed Bougara, Boumerdes
Faculté des Sciences de l'Ingénieur
Département de Génie Electrique et Electronique



Continuous Speech Recognition

Thesis
presented in partial fulfilment of the requirements of the degree of
Docteur es Sciences
in
Electronic System Engineering
by
Abdelhakim DAHIMENE

Before the jury composed of:

Dr. Aksas R. (Professeur, ENP)	President
Dr. Guessoum A. (Professeur, Université S. Dahlab, Blida)	Examiner
Dr. Trabelsi M. (Maître de conférences, ENP)	Examiner
Dr. Hariche K. (Professeur, UMBB)	Examiner
Dr. Noureddine (Maître de conférences, UMBB)	Supervisor

April 2009

ملخص

يعد تمييز الكلام أحد أهم تحديات بداية هذا القرن. كما سمحت التطورات المذهلة للأجهزة الإلكترونية باستعمال طرق رياضية غاية في التعقيد محل مشاكل جد مستعصية. في هذه المذكرة، سنحاول أن نبرهن أن الطرق التسلسلية المخترعة لفك الرموز "convolutionnels" ذات العوائق جد الطويلة يمكن أن تستعمل في تمييز الكلام.

أهم المساهمات في هذه المذكرة يمكن تلخيصها فيما يلي:

1. تطبيق خوارزمية البطارية (التسلسلية) لتمييز الكلام المتواصل.
2. تطوير و تحليل مقياس للمسافة (داخل شجرة) بالاعتماد على مقياس المسافة لهماهالانوبيس. هذا المقياس استخدم لبرمجة خوارزمية البطارية في برنامج التمييز و كذا في خوارزمية فيتربي في برنامج التلقين.
3. تطوير خوارزمية جديدة بناء على التوقع الخطي لاسترجاع إشارة الكلام المبتورة.
4. تطوير طريقة لإزالة الضوضاء في إشارة الكلام بناء على مصفاة وينر ذات درجة منخفضة.
5. للحصول على مقياس مقبول للمسافة، قمنا بتحليل إحصائي لثلاث تمثيلات وسيطية للكلام و أثبتنا أن معاملات (MFCC) تتبع توزيع دالة غوس و تعطي أفضل بين الأقسام مقارنة بمعاملات (LPC) و (PARCOR).
6. في آخر فصل من المذكرة، قمنا بتطوير خوارزمية آلية للتجزئة والتي نستعملها في تلقين برامج تمييز الكلام.

Résumé

La reconnaissance de la parole continue est un des plus grands défis dans ce début du siècle. Les avancées impressionnantes dans l'équipement électronique permettent l'utilisation de méthodes mathématiques sophistiquées pour résoudre des problèmes extrêmement complexes. Dans cette thèse, nous allons montrer que les méthodes séquentielles inventées pour décoder les codes convolutionnels de grande longueur de contrainte peuvent être utilisées dans la reconnaissance de la parole.

Les contributions principales de cette thèse peuvent être récapitulées comme suit:

(1) L'applicabilité de l'algorithme de décodage à pile (séquentiel) à la reconnaissance de la parole continue.

(2) Le développement et l'analyse d'une métrique de chemin (dans un arbre) basé sur la distance de Mahalanobis. Cette métrique a été utilisée dans la mise en place de l'algorithme de pile dans le programme de reconnaissance et également dans l'algorithme de Viterbi dans le programme d'apprentissage.

(3) Le développement d'un nouvel algorithme basé sur la prédiction linéaire pour la restauration d'un signal de parole échantillé.

(4) Une méthode de suppression du bruit dans le signal de parole basée sur des filtres de Wiener à paramètres non stationnaires. Nous avons obtenu des résultats remarquables à l'aide d'un filtre d'ordre très réduit.

(5) Afin de trouver une bonne métrique de chemin pour notre algorithme à pile, nous avons effectué une analyse statistique de trois représentations paramétriques de la parole et nous avons prouvé que les paramètres de MFCC sont pratiquement Gaussien et fournissent la meilleure séparabilité entre classes par rapport aux coefficients de LPC et de PARCOR.

(6) Dans le dernier chapitre de la thèse, nous avons développé un algorithme automatique de segmentation que nous utilisons pour l'apprentissage du programme de reconnaissance de la parole.

Abstract

Continuous speech recognition is one of the greatest challenges in this beginning of the century. The impressive advances in hardware allow the use of sophisticated mathematical methods to solve complex problems. In this thesis, we show that methods invented for solving long constraint length convolutional codes can be used in speech recognition.

The main contributions of this thesis can be summarized as follows:

(1) The applicability of the stack decoding algorithm to continuous speech recognition.

(2) The development and the analysis of a path metric based on the Mahalanobis distance. This path metric has been used in the implementation of the stack algorithm in recognition program and also in the Viterbi algorithm in the training program.

(3) The development of a novel algorithm for clipped speech restoration based on linear prediction.

(4) Speech denoising method based on time varying Wiener filters. We obtained remarkable results using a very low order filter.

(5) In order to develop a good path metric for our stack algorithm, we have performed a statistical analysis of three parametric representations of speech and we have shown that the MFCC set is nearly Gaussian and provides the best separability between classes as compared with LPC and PARCOR coefficients.

(6) In the last chapter of the thesis, we have developed an automatic segmentation algorithm that we use for training the speech recognition program.

Table of Contents

Acknowledgments	ii
ملخص	iii
Résumé	iv
Abstract	v
Table of Contents	vi
List of Illustrations	x
List of Tables	xiv
Abbreviations	xv
Chapter 1	1
Introduction	1
Chapter 2	5
The Speech Signal	5
2.1 A Production Model of the Speech Signal	5
2.1.1 Vocal tract model	7
2.1.2 Radiation Model	8
2.1.3 Glottal Pulse Model	8
2.1.4 The Complete Model	8
2.2 Short Time Spectrum Analysis	9
2.2.1 Definition	9
2.2.2 The Sound Spectrogram	11
2.3 Phonetics and Phonology	13
Vowels	13
Semivowels	14
Nasals	15

Unvoiced Fricatives	15
Voiced Fricatives	15
Voiced Stops (Plosives)	15
Unvoiced Stops (Plosives)	15
2.4 Short Time Characterization of Speech	16
Chapter 3	20
Clipped Speech Restoration	20
3.1 Basic Interpolation Methods	20
3.2 Justification of the Method	21
3.3 The Proposed Restoration Algorithm	22
3.3.1 Computation of the prediction coefficients	23
3.3.2 Interpolation of the missing samples	24
3.4 Results	24
3.4.1 Synthetic Speech	25
3.4.2 Artificially Clipped Natural Speech	29
3.4.3 Clipped Natural Speech	30
3.4.4 Discussion	32
Chapter 4	34
Speech Denoising using Wiener Filters	34
4.1 Wiener Filter Theory	34
4.1.1 Non Causal IIR Wiener Filters	36
4.1.2 Causal IIR Wiener Filter	37
4.1.3 Causal FIR Wiener Filter	38
4.2 The Autoregressive Model	38
4.3 Time Invariant Wiener Filtering	40
4.3.1 FIR Time Invariant Wiener Filter	40
4.3.2 IIR Causal Wiener Filter	42

4.3.4.a First Order Model	42
4.3.4.b Second Order Model	44
4.3.3 IIR Non Causal Wiener Filter	46
4.3.3.a First Order Model	46
4.3.3.b Second Order Model	48
4.4 Time Varying Wiener Filter	50
Chapter 5	57
Feature Extraction and Selection	57
5.1 Linear Prediction	57
5.1.1 The Autocorrelation Formulation	57
5.1.2 The Covariance Formulation	60
5.2 The Mel Frequency Cepstral Coefficients	62
5.3 Statistical Comparison of Features	64
5.3.1 The Normal Density	64
5.3.2 Assessing the Assumption of Normality	65
5.3.3 Discriminant Analysis	70
Chapter 6	74
Continuous Speech Recognition	74
6.1 Statistical Pattern Recognition	74
6.2 Finite State Automata.....	76
Finite State Automaton Definition [53]:	76
6.3 Hidden Markov Models	78
6.4 Benchmark System [76].....	82
6.5 Sequential Decoding Applied to Speech Recognition [20]	86
6.5.1 Speech Production Model	87
6.5.2 The Stack Algorithm	91
6.5.3 Implementation of the Algorithm	95

Chapter 7	98
Automatic Training	98
7.1 Finite State Automaton Update	98
7.2 Speech Segmentation	99
7.2.1 Clustering Methods	99
7.2.1.a Hierarchical Clustering	99
7.2.1.b Non-hierarchical Clustering	101
7.2.2 Clustering results	102
7.2.3 Segmentation using the Viterbi Algorithm	105
Chapter 8	111
Conclusion	111
Suggestion for Further Research	112
References	113
Appendix A	121
The Adopted List of Phonemes	121