CONTENTS

Abstract **vii**

Contents **ix**

1 Introduction **1**
1.1 Motivation and document structure 2
1.2 Scientific and technological goals 4
1.3 Context of this thesis 5
1.3.1 EMMA MOOCs 6
1.3.2 The UPV media repository 9
Bibliography 11

2 Preliminaries **15**
2.1 Introduction 16
2.2 Language modelling 16
2.3 Automatic speech recognition 17
2.3.1 Acoustic modelling 18
2.3.2 Lexical modelling 20
2.3.3 The recognition process 21
2.4 Machine translation 22
2.4.1 Translation modelling 22
2.4.2 The decoding process 24
2.5 Evaluation measures 24
Bibliography 25

3 Automatic transcription of Spanish video lectures **29**
3.1 Introduction . . . 30
3.2 The poliMedia corpus . . . 30
3.3 Experiments on speaker and topic adaptation . . . 30
3.4 Integration of ASR into the Matterhorn platform . . . 33
3.5 Conclusions . . . 36
Bibliography . . . 36

4 User evaluations on interaction in the review of video lectures **39**
4.1 Introduction . . . 40
4.2 Methodology of the user trials . . . 41
4.3 Experimental results . . . 43
4.3.1 First phase: Post-editing . . . 43
4.3.2 Second phase: Intelligent interaction . . . 47
4.3.3 Third phase: Two-step supervision . . . 50
4.4 Conclusions . . . 52
Bibliography . . . 54

5 Automatic transcription and translation systems for MOOCs and OER **57**
5.1 Introduction . . . 58
5.2 Transcription systems . . . 58
5.2.1 Italian . . . 58
5.2.2 Portuguese . . . 61
5.3 Translation systems . . . 64
5.3.1 English language model . . . 65
5.3.2 Italian-English . . . 66
5.3.3 Portuguese-English . . . 68
5.3.4 Dutch-English . . . 70
5.3.5 English-Italian . . . 72
5.4 Conclusions . . . 73
Bibliography . . . 75

6 Evaluation on the review of multilingual videos for MOOCs and OER **77**
6.1 Introduction . . . 78
6.2 Integration of ASR and MT systems . . . 78
6.3 Transcription and translation quality . . . 80
6.3.1 Transcription quality . . . 81
6.3.2 Translation quality . . . 82

6.3.3 Comparison with mainstream providers 83
6.4 Reviewing time 84
6.4.1 Transcription reviewing time 85
6.4.2 Translation reviewing time 87
6.4.3 Reviewing time across languages 88
6.5 Impact on the case studies 89
6.5.1 The EMMA platform 89
6.5.2 The UPV media repository 92
6.6 Conclusions 92
Bibliography 93

7 Conclusions **95**
7.1 Summary and future work 96
7.2 Contributions 97
Bibliography 100

List of figures **103**

List of tables **105**

List of abbreviations **107**