
CONTENTS

Agraiments	VII
Resum	IX
Resumen	XI
Abstract	XIII
List of Acronyms	XVIII
List of Mathematical Symbols	XIX
List of Figures	XXII
List of Tables	XXIV
I Introduction	1
1 Introduction	3
1.1 Motivation	3
1.2 Scientific and Technological Objectives	4
1.3 Document Structure	5
Bibliography	5
2 Background	7
2.1 Statistical Natural Language Recognition	8
2.2 Preprocessing and Feature Extraction	9
2.2.1 Automatic Speech Recognition Features	9
2.2.2 <i>Off-line</i> Handwriting Text Recognition Features	10
2.2.3 <i>On-line</i> Handwriting Text Recognition Features	10
2.2.4 Tandem Features	11
2.3 Statistical Modelling	11
2.3.1 Morphological Modelling	12
2.3.2 Language Modelling	17

2.3.3	Lexicon Modelling	19
2.4	Decoding	20
2.4.1	The Viterbi Algorithm	20
2.4.2	Recognition Output Formats	22
2.5	Assistive Transcription of Historical Manuscripts	24
2.6	Crowdsourcing for Natural Language Processing Tasks	25
2.7	Evaluation Measures	25
2.7.1	Natural Language Recognition Evaluation	25
2.7.2	Language Model Evaluation	26
2.7.3	Computer Assisted Transcription Evaluation	27
2.7.4	Multimodal Crowdsourcing	27
2.7.5	Statistical Significance	28
2.8	Datasets	28
2.8.1	Historical Manuscript Corpora (<i>Off-line</i> Handwriting)	29
2.8.2	Touch Screen Handwriting Corpus (<i>On-line</i> Handwriting): UNIPEN	29
2.8.3	Training Speech Corpus: <i>Albayzin</i>	31
2.8.4	Multimodal (Text - Speech) Corpora	31
	Bibliography	32
 II Multimodality		 41
 3 Combining Handwriting and Speech		 43
3.1	Introduction	43
3.2	Hypothesis Combination on Natural Language Recognition	44
3.2.1	Recogniser Output Voting Error Reduction (ROVER)	45
3.2.2	N-best ROVER	45
3.2.3	Lattices Rescoring	46
3.3	Our proposal: Bimodal Confusion Network Combination	46
3.3.1	Subnetworks Based Alignment	47
3.3.2	Composing a New Confusion Network	47
3.4	Conclusions	49
	Bibliography	49
 4 Multimodal Experimental Results		 53
4.1	Experimental Framework	54
4.1.1	Datasets	54
4.1.2	Features	55
4.1.3	Models	55

4.1.4	Evaluation Metrics	56
4.2	Experimental Setup	56
4.3	Experiment 1: Iterative and Non-Iterative Combination	57
4.3.1	Experiments with <i>Cristo Salvador</i>	57
4.3.2	Experiments with <i>Rodrigo</i>	57
4.4	Experiment 2: Unimodal and Multimodal Combination	58
4.4.1	Baseline Experiments	58
4.4.2	Unimodal Combination Experiments	59
4.4.3	Multimodal Combination Experiment	60
4.4.4	Difficulty of Reaching the Oracle Values	60
4.5	Experiment 3: Multimodal Combination Comparative	61
4.6	Conclusions and Future Work	62
	Bibliography	63
III	Interactivity	65
5	Assistive Transcription	67
5.1	Computer Assisted Transcription Overview	68
5.2	Multimodal Computer Assisted Transcription	69
5.2.1	Multimodal Hypotheses Combination in CATTI	70
5.2.2	Multimodal Hypotheses Correction in CATTI	71
5.3	Conclusions	73
	Bibliography	73
6	Interactivity Experimental Results	75
6.1	Experimental Framework	76
6.1.1	Datasets	76
6.1.2	Features	76
6.1.3	Models	77
6.1.4	Evaluation Metrics	77
6.1.5	Experimental Setup	77
6.2	Experiment 1: Multimodal Hypotheses Combination	78
6.2.1	Experiments with <i>Cristo Salvador</i>	78
6.2.2	Experiments with <i>Rodrigo</i>	79
6.3	Experiment 2: Multimodal Hypotheses Correction	80
6.3.1	<i>Off-line</i> and <i>On-line</i> HTR Results	80
6.3.2	CATTI and Multimodal CATTI Results	80
6.4	Experiment 3: Multimodal Hypotheses Combination and Correction	81

6.4.1	Post-Edition Baseline Results	81
6.4.2	CATTI Results	82
6.4.3	Multimodal CATTI Results	82
6.5	Conclusions and Future Work	83
	Bibliography	84
IV	Crowdsourcing	85
7	Collective Collaboration	87
7.1	Multimodal Crowdsourcing Framework	88
7.1.1	Language Model Interpolation	89
7.1.2	Multimodal Combination	90
7.1.3	Reliability Verification	90
7.1.4	Lines Selection	91
7.1.5	Client Application for Speech Acquisition	91
7.2	Conclusions	92
	Bibliography	92
8	Crowdsourcing Experiments	95
8.1	Experimental Conditions	96
8.1.1	Datasets	96
8.1.2	Features	97
8.1.3	Models	97
8.1.4	Evaluation Metrics	98
8.1.5	Experimental Setup	98
8.2	Experiment 1: Supervised Multimodal Crowdsourcing	98
8.2.1	Baseline and Framework Adjustment	98
8.2.2	Speaker Ordering	99
8.2.3	ASR Reliability Verification	100
8.2.4	Absence of Speech Utterances	101
8.2.5	Collaborator Effort Optimisation	101
8.3	Experiment 2: Unsupervised Multimodal Crowdsourcing	103
8.3.1	Baseline and Framework Adjustment	103
8.3.2	Preliminary Experiments	103
8.3.3	ASR Reliability Verification and Collaboration Effort	104
8.3.4	Collaboration Effort per Line	106
8.4	Conclusions and Future Work	109
	Bibliography	109

V	Conclusions and Future Work	111
9	Conclusions and Future Work	113
9.1	Conclusions	113
9.2	Scientific Work and Contributions	114
9.3	Future Work	117
	Bibliography	118