

Table of Contents

Abstract	iii
Acknowledgements	v
1 Introduction.....	1
1.1 Speech Recognition	2
1.2 Neural Networks.....	4
1.3 Thesis Outline.....	7
2 Review of Speech Recognition.....	9
2.1 Fundamentals of Speech Recognition	9
2.2 Dynamic Time Warping.....	14
2.3 Hidden Markov Models	15
2.3.1 Basic Concepts.....	16
2.3.2 Algorithms	17
2.3.3 Variations	22
2.3.4 Limitations of HMMs.....	26
3 Review of Neural Networks	27
3.1 Historical Development	27
3.2 Fundamentals of Neural Networks.....	28
3.2.1 Processing Units.....	28
3.2.2 Connections	29
3.2.3 Computation.....	30
3.2.4 Training	35
3.3 A Taxonomy of Neural Networks	36
3.3.1 Supervised Learning.....	37
3.3.2 Semi-Supervised Learning.....	40
3.3.3 Unsupervised Learning.....	41
3.3.4 Hybrid Networks	43
3.3.5 Dynamic Networks.....	43
3.4 Backpropagation.....	44
3.5 Relation to Statistics.....	48

4	Related Research	51
4.1	Early Neural Network Approaches	51
4.1.1	Phoneme Classification	52
4.1.2	Word Classification	55
4.2	The Problem of Temporal Structure	56
4.3	NN-HMM Hybrids	57
4.3.1	NN Implementations of HMMs	57
4.3.2	Frame Level Training	58
4.3.3	Segment Level Training	60
4.3.4	Word Level Training	61
4.3.5	Global Optimization	62
4.3.6	Context Dependence	63
4.3.7	Speaker Independence	66
4.3.8	Word Spotting	69
4.4	Summary	71
5	Databases	73
5.1	Japanese Isolated Words	73
5.2	Conference Registration	74
5.3	Resource Management	75
6	Predictive Networks	77
6.1	Motivation... and Hindsight	78
6.2	Related Work	79
6.3	Linked Predictive Neural Networks	81
6.3.1	Basic Operation	81
6.3.2	Training the LPNN	82
6.3.3	Isolated Word Recognition Experiments	84
6.3.4	Continuous Speech Recognition Experiments	86
6.3.5	Comparison with HMMs	88
6.4	Extensions	89
6.4.1	Hidden Control Neural Network	89
6.4.2	Context Dependent Phoneme Models	92
6.4.3	Function Word Models	94
6.5	Weaknesses of Predictive Networks	94
6.5.1	Lack of Discrimination	94
6.5.2	Inconsistency	98

7	Classification Networks	101
7.1	Overview	101
7.2	Theory	103
7.2.1	The MLP as a Posterior Estimator	103
7.2.2	Likelihoods vs. Posteriors	105
7.3	Frame Level Training	106
7.3.1	Network Architectures	106
7.3.2	Input Representations	115
7.3.3	Speech Models	119
7.3.4	Training Procedures	120
7.3.5	Testing Procedures	132
7.3.6	Generalization	137
7.4	Word Level Training	138
7.4.1	Multi-State Time Delay Neural Network	138
7.4.2	Experimental Results	141
7.5	Summary	143
8	Comparisons	147
8.1	Conference Registration Database	147
8.2	Resource Management Database	148
9	Conclusions	151
9.1	Neural Networks as Acoustic Models	151
9.2	Summary of Experiments	152
9.3	Advantages of NN-HMM hybrids	153
Appendix A. Final System Design		155
Appendix B. Proof that Classifier Networks Estimate Posterior Probabilities		157
Bibliography		159
Author Index		169
Subject Index		173