

# Inhaltsverzeichnis

<b>1 Einleitung</b> .....	1
<b>2 Grundlagen zur Spracherkennung mit stochastischen Modellen</b> .....	5
2.1 Aufbereitung des Sprachsignals .....	5
2.2 Modellstrukturen .....	6
2.2.1 Multimodale Hidden-Markov-Modelle .....	6
2.2.2 Stochastische Markov-Graphen .....	7
2.3 Basisfunktionen .....	8
2.3.1 Unspezifische Basisfunktionen .....	8
2.3.2 Phonemspezifische Basisfunktionen .....	9
2.3.3 Zustandsspezifische Basisfunktionen .....	10
2.4 Initialisierung stochastischer Modelle .....	11
2.5 Training stochastischer Modelle .....	11
2.6 Erkennung fließender Sprache .....	12
<b>3 Erzeugung, Training und Evaluierung von Hidden-Markov-Modellen</b> .....	15
3.1 Generierung der Basisfunktionen .....	15
3.1.1 Erweiterter LBG-Algorithmus .....	15
3.1.2 Abstandsmaße für die Clustervereinigung .....	18
3.1.2.1 Messung des Emissionsverhältnisses .....	18
3.1.2.2 Messung der Veränderung der Intra-Cluster-Eigenwerte .....	19
3.1.3 Vereinigung zweier Cluster .....	20
3.2 Initialisierung von Hidden-Markov-Modellen .....	21
3.3 Maximum-Likelihood-Training stochastischer Modelle .....	23
3.3.1 Allgemeine Formulierung .....	23
3.3.2 Viterbi-Training .....	25
3.3.2.1 Initialisierung der temporären Parameter .....	25
3.3.2.2 Viterbi mit automatischer Segmentierung und Backtracking .....	26
3.3.2.3 Parameternormalisierung und -aktualisierung .....	27
3.3.3 Forward-Backward-Training .....	28
3.3.4 Anmerkungen zu den Trainingsverfahren .....	32
3.4 Phonemererkennung mit stochastischen Modellen .....	33
3.4.1 Freilaufende Phonemererkennung mit dem Viterbi-Algorithmus .....	34
3.4.2 Phonemererkennung mit dem Forward-Backward-Algorithmus .....	35
3.4.3 Beschleunigung der Emissionsberechnung .....	36
3.5 Experimentelle Betrachtung und Interpretation .....	37

## INHALTSVERZEICHNIS

3.5.1 Vorverarbeitung und Sprachdatenbasis .....	37
3.5.2 Basisfunktionsgenerierung .....	38
3.5.3 Evaluierung von Hidden-Markov-Modellen bei gegebener Segmentierung .....	40
3.5.3.1 Vergleich zwischen Viterbi- und Forward-Backward-Verarbeitung .....	40
3.5.3.2 Vergleich zwischen HMMs mit unspezifischen und phonem-spezifischen Basisfunktionen .....	41
3.5.4 Freilaufende Evaluierung von Hidden-Markov-Modellen .....	42
3.5.4.1 Evaluierung ohne Ableitungsmerkmale .....	42
3.5.4.2 Evaluierung mit Ableitungsmerkmalen .....	45
<b>4 Stochastische Markov-Graphen .....</b>	<b>49</b>
4.1 Einführung und Motivation .....	49
4.2 Erzeugung stochastischer Markov-Graphen .....	51
4.2.1 Ableitung von Hidden-Markov-Modellen .....	51
4.2.2 Allgemeine Vorgehensweise .....	53
4.2.3 Generierung kontextabhängiger Basisfunktionen .....	53
4.2.4 Modelltraining .....	56
4.3 Perplexität stochastischer Markov-Graphen .....	56
4.4 Ergebnisse und Interpretation .....	59
4.4.1 Viterbi- und Forward-Backward-Abarbeitung .....	59
4.4.2 Veranschaulichung der entstehenden Graphenstruktur .....	60
4.4.3 Evaluierung stochastischer Markov-Graphen .....	62
4.4.3.1 Evaluierung von Modellen ohne Ableitungsmerkmale .....	63
4.4.3.2 Evaluierung von Modellen mit Ableitungsmerkmalen .....	66
4.4.3.3 Transitionsgewichtung .....	68
4.4.3.4 Perplexitätsmessungen .....	70
4.5 Vergleich mit anderen Ansätzen zur Trajektorienmodellierung .....	71
4.6 Diskussion .....	74
<b>5 Explizite Zustandverweildauermodellierung .....</b>	<b>79</b>
5.1 Motivation .....	79
5.2 Viterbi-Abarbeitung .....	80
5.3 Verteilungsdichtefunktion .....	81
5.4 Evaluierung und Interpretation .....	82
<b>6 Diskriminatives Training .....</b>	<b>85</b>
6.1 Motivation .....	85
6.2 GPD-Training bei fester Phonemsegmentierung .....	86
6.2.1 Formalismus zum Training der Standard-Parameter .....	86
6.2.2 Training diskriminativer Zustandsgewichte .....	91

## INHALTSVERZEICHNIS

6.2.2.1 Verrechnung ohne Verzerrung des Viterbi-Pfades .....	92
6.2.2.2 Verrechnung mit Verzerrung des Viterbi-Pfades .....	95
6.3 GPD-Training mit impliziter Segmentierung .....	96
6.4 Evaluierung und Interpretation .....	99
6.4.1 Zustandsgewichtung .....	99
6.4.1.1 Evaluierungsergebnisse .....	99
6.4.1.2 Beispielhafte Veranschaulichung der resultierenden Zustandsgewichte .	101
6.4.2 Diskr. Training aller Modellparameter in fester Phonemsegmentierung .....	103
6.4.3 Diskr. Training aller Modellparameter mit impliziter Segmentierung .....	105
6.5 Diskussion .....	106
<b>7 Erkennung fließender Sprache auf Wortebene .....</b>	<b>111</b>
7.1 Viterbi-Beam-Search-Algorithmus .....	111
7.1.1 Motivation und Prinzip .....	111
7.1.2 Repräsentation der Informationsquellen .....	113
7.1.2.1 Akustische Modelle .....	113
7.1.2.2 Lexikon .....	114
7.1.2.3 Sprachmodell .....	115
7.1.3 Hypothesenmanagement .....	116
7.1.3.1 Zustandshypothesen .....	116
7.1.3.2 Buchführung zur lokalen Rekombination .....	118
7.1.3.3 Ablegen der Backtrackinginformation .....	121
7.1.3.4 Pruning der Zustandshypothesen .....	123
7.1.4 Erweiterung für Zwischenwortaussprachevarianten .....	124
7.1.5 Einsatz zur Segmentierung .....	125
7.2 Experimentelle Ergebnisse und Interpretation .....	126
7.2.1 Evaluierung von HMMs und SMGs mit Standardlexika .....	126
7.2.1.1 Evaluierung auf der Diphon-Datenbasis .....	127
7.2.1.2 Evaluierung auf der Erba-Datenbasis .....	130
7.2.1.3 Evaluierung auf der Verbmobil-Datenbasis .....	132
7.2.2 Evaluierung von SMGs mit Aussprachevarianten .....	133
7.3 Diskussion .....	135
<b>8 Gesamtdiskussion .....</b>	<b>139</b>
<b>Literaturverzeichnis .....</b>	<b>143</b>
<b>Verzeichnis der wichtigsten Symbole .....</b>	<b>157</b>