

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	Einleitung . . . . .	1
1.2	Was macht Sprache so schwierig? . . . . .	2
1.3	Literatur . . . . .	4
<b>2</b>	<b>Sprachverarbeitung</b>	<b>5</b>
2.1	Spracherzeugung . . . . .	5
2.1.1	Quelle-Filter-Modell . . . . .	8
2.2	Einheiten . . . . .	8
2.3	Variabilität . . . . .	11
2.4	Übungen . . . . .	13
<b>3</b>	<b>Spracherkennung</b>	<b>15</b>
3.1	Techniken und Anwendungen . . . . .	15
3.2	Schwierigkeitsstufen . . . . .	18
3.3	Aufbau eines komplexen Systems . . . . .	19
3.4	Fehlerarten . . . . .	21
3.4.1	Spracherkennung . . . . .	21
3.4.2	Sprecherverifikation . . . . .	23
3.5	Übungen . . . . .	27
<b>4</b>	<b>Vorverarbeitung</b>	<b>29</b>
4.1	Sprachdetektion . . . . .	29
4.2	Merkmalsextraktion . . . . .	30
4.2.1	Blockbildung . . . . .	30
4.2.2	Diskrete Fourier-Transformation . . . . .	32
4.2.3	Anpassung des Frequenzbereiches . . . . .	34
4.2.4	Lineare Prädiktion . . . . .	36
4.2.5	Cepstrale Darstellung . . . . .	43
4.2.6	Dynamische Merkmale . . . . .	45
4.2.7	Andere Merkmale . . . . .	47
4.3	Vektor-Quantisierung . . . . .	48
4.4	Übungen . . . . .	51

<b>5</b>	<b>Mustervergleich</b>	<b>53</b>
5.1	Grundprinzip . . . . .	53
5.2	Dynamische Zeitverzerrung . . . . .	55
5.3	Suche . . . . .	57
5.3.1	Aufwand . . . . .	63
5.4	Bewertung . . . . .	64
5.5	Übungen . . . . .	65
<b>6</b>	<b>Hidden-Markov-Modelle</b>	<b>67</b>
6.1	Einleitung . . . . .	67
6.2	Ein anschauliches Beispiel . . . . .	67
6.2.1	Übertragung auf Sprachsignale . . . . .	69
6.3	Mathematische Grundlagen . . . . .	70
6.3.1	Wahrscheinlichkeiten . . . . .	70
6.3.2	Grundstruktur von HMM . . . . .	71
6.4	Ausgabesymbole . . . . .	74
6.5	Anwendung . . . . .	75
6.5.1	Viterbi-Algorithmus . . . . .	79
6.5.2	Modelltraining . . . . .	82
6.5.3	Kontinuierliche Symbole . . . . .	85
6.6	Diskriminatives Modelltraining . . . . .	89
6.6.1	Merkmalstransformation . . . . .	94
6.7	Vergleich zwischen HMM und DTW . . . . .	96
6.8	Übungen . . . . .	97
<b>7</b>	<b>Einsatz von Hidden-Markov-Modellen</b>	<b>99</b>
7.1	Einleitung . . . . .	99
7.2	Ganzwortmodelle . . . . .	99
7.3	Phonemmodelle . . . . .	101
7.3.1	Aussprachelexikon . . . . .	101
7.3.2	Kombination zu Wortmodellen . . . . .	103
7.3.3	Kontextabhängige Modelle . . . . .	104
7.3.4	Training und Erkennung . . . . .	107
7.4	Datenbanken . . . . .	108
7.5	Kontinuierliche Erkennung . . . . .	109
7.5.1	Netzwerke . . . . .	109
7.5.2	Statistische Sprachmodelle . . . . .	112
7.5.3	Effiziente Suche . . . . .	118
7.6	Schnittstelle zu anderen Modulen . . . . .	118
7.6.1	N-Besten Wortfolgen . . . . .	118
7.6.2	Wordhypothesengraph . . . . .	120
7.7	Übungen . . . . .	121

<b>8</b>	<b>Syntax</b>	<b>123</b>
8.1	Einleitung . . . . .	123
8.2	Grundlagen . . . . .	123
8.3	Grammatiken . . . . .	125
8.3.1	Einseitig lineare Grammatiken . . . . .	127
8.3.2	Kontextfreie Grammatiken . . . . .	128
8.3.3	Wortproblem . . . . .	129
8.4	Grammatiken für Sprachanwendungen . . . . .	130
8.4.1	Java Speech Grammar Format . . . . .	130
8.4.2	Speech Recognition Grammar Specification . . . . .	131
8.4.3	HTK-Grammatik . . . . .	133
8.5	Merkmalsstrukturen . . . . .	133
8.6	Übungen . . . . .	135
<b>9</b>	<b>HTK</b>	<b>137</b>
9.1	Sprachdaten und Vorbereitungen . . . . .	138
9.2	Merkmalsextraktion . . . . .	140
9.3	Ganzwortmodelle . . . . .	142
9.3.1	Initialisierung . . . . .	142
9.3.2	Einzelne Wörter . . . . .	143
9.3.3	Wortketten . . . . .	147
9.4	Phonemmodelle . . . . .	147
9.5	Übungen . . . . .	151
<b>10</b>	<b>Dialogsysteme</b>	<b>153</b>
10.1	Einleitung . . . . .	153
10.2	Dialogtypen . . . . .	153
10.3	Programmierschnittstellen . . . . .	156
10.4	Beschreibungssprachen . . . . .	157
10.5	Speech Application Language Tags (SALT) . . . . .	157
10.6	Voice Extensible Markup Language . . . . .	159
10.6.1	Grundelemente . . . . .	159
10.6.2	Ereignisse . . . . .	163
10.6.3	Gemischte Initiative . . . . .	164
10.7	Übungen . . . . .	166
<b>11</b>	<b>Friedberger Java-Sprach-Tools</b>	<b>167</b>
11.1	Einleitung . . . . .	167
11.2	DTW-Erkenner fbdtw . . . . .	167
11.3	FBVIEW . . . . .	169
11.3.1	Dateiformate . . . . .	171
11.3.2	Signalanalyse . . . . .	172
11.3.3	Transkriptionen . . . . .	172

11.3.4 TCP-Schnittstelle . . . . .	173
11.4 FBGenerator . . . . .	173
11.5 VoiceXML Interpreter . . . . .	174
<b>A Hidden-Markov-Modelle</b>	<b>175</b>
A.1 Baum-Welch-Algorithmus . . . . .	175
A.2 Skalierung . . . . .	179
<b>B Abkürzungen</b>	<b>183</b>
<b>Literaturverzeichnis</b>	<b>185</b>
<b>Index</b>	<b>198</b>