

Dipl.-Ing. Lars Knohl, Schermbeck

**Prosodiegesteuerte
Sprecher- und Umwelt-
adaption in einer
Mehrsprecher-Architektur
zur maschinellen Sprach-
erkennung**

Reihe **10**: Informatik/
Kommunikationstechnik Nr. **461**

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen der Sprecheradaption	4
2.1	Sprechervariabilität	6
2.1.1	Sprecherspezifische Variabilität	7
2.1.2	Umweltspezifische Variabilität	10
2.2	Konzepte der Spracherkennung	11
2.2.1	Sprecheradaptive Spracherkennung	14
2.2.2	Technische Komponenten der Spracherkennung	17
2.3	Konzepte der Sprecheradaption	21
2.3.1	Textunabhängige (unüberwachte) Sprecheradaption	24
2.3.1.1	Spektrale Subtraktion	24
2.3.1.2	Kanal-Kompensation (Channel Equalization)	25
2.3.1.3	Frequenzachsentransformation	27
2.3.1.4	Kodebuchadaption	29
2.3.2	Textabhängige (überwachte) Sprecheradaption	32
2.3.2.1	Frequenzachsentransformation	32
2.3.2.2	Kodebuchadaption	41
2.3.2.3	HMM-Adaption	50
2.4	Zusammenfassung	53
3	Prosodiegesteuerte Sprecher- und Umweltadaption	56
3.1	Nicht-lineare Zeitachsentransformation in der textabhängigen Sprecheradaption	57
3.2	Das Adaptionsverfahren	63
3.2.1	Extraktion der Log-F1-Energie als prosodisches Hilfsmerkmal	68

3.2.2	Segmentierung des Log-F1-Energieverlaufes	71
3.3	Experimentelle Untersuchungen	74
3.3.0.1	Die Testumgebung	75
3.3.0.2	Das Testkorpus	77
3.3.1	Optimierung der Verfahrensparameter	80
3.3.2	Performanz in störungsarmer Sprechumgebung	86
3.3.3	Performanz in störungsbehafteter Sprechumgebung	88
3.3.3.1	Trainingsfremde Sprechumgebung	88
3.3.3.2	Sprecheradaption vs. sprecherunabhängiger und sprecher- abhängiger Erkennung in trainingsgetreuer Sprechumgebung	94
3.4	Zusammenfassung	96
4	Sprecher- und Umweltadaption in einer Mehrsprecher-Architektur	99
4.1	Die Mehrsprecher-Architektur	103
4.1.1	Methoden der Kodebuch-Selektion	105
4.2	Experimentelle Untersuchungen	108
4.2.0.1	Das Testkorpus	109
4.2.1	Bestimmung des Selektionsverfahrens und einer Identifikationsäußerung	111
4.2.2	Performanz in störungsarmer Sprechumgebung	113
4.2.3	Performanz in störungsbehafteter Sprechumgebung	115
4.2.3.1	Trainingsfremde Sprechumgebung	115
4.2.3.2	Sprecheradaption vs. sprecherunabhängiger und sprecher- abhängiger Erkennung in trainingsgetreuer Sprechumgebung	119
4.3	Zusammenfassung	120
5	Zusammenfassung	123
A	Die Testumgebung	130
A.1	Der Preprozessor	131
A.2	Der Vektorquantisierer	133
A.2.1	Vektorquantisierung mit kohonschen Merkmalskarten	134

A.3 Der Erkennen	138
B Das Testkorpus	141
C Sprechumgebungen	144
C.1 Umgebungsrauschen	145
C.1.1 SUN-SPARC10 Lüfterrauschen	148
C.2 Übertragungsstrecken	149
C.2.1 Raum	149
C.2.2 Telefon	153
C.3 Kombinierte Szenarien	155
C.3.1 Telefon + Leitungsrauschen	156
C.3.2 Lüfterrauschen + Raum + Telefon + Leitungsrauschen	158
D Erkennungsergebnisse	161
E Literaturverzeichnis	167
F Stichwortverzeichnis	179