

Inhaltsangabe

In der automatischen Spracherkennung hat sich seit Mitte der 80er Jahre ein statistischer Modellierungsansatz auf der Basis von Markovmodellen (HMMs) als Standard etabliert. Ein wesentlicher Schwachpunkt dieses Ansatzes besteht in der fehlenden Ausnutzung statistischer Abhängigkeiten zwischen den Merkmalsvektoren aufeinander folgender Sprachrahmen. Dies betrifft sowohl die akustische Modellierung mit HMMs im Back-End des Erkenners als auch eine vorangehende modellbasierte Merkmalsentstörung auf der Basis von Gaußmischungsmodellen (GMMs). An dieser Stelle setzt die vorliegende Arbeit an, deren Ziel die Berücksichtigung statistischer Abhängigkeiten zwischen den Sprachmerkmalen verschiedener Sprachrahmen innerhalb des statistischen Spracherkennungsansatzes ist. Bei der Merkmalsentstörung können Inter-Frame Korrelationen ausgenutzt werden, indem die Dynamik der cepstralen Sprachmerkmale mit schaltenden, linearen Dynamikmodellen (SLDMs) beschrieben wird. Es werden verschiedene Möglichkeiten untersucht, um die a posteriori Wahrscheinlichkeit der Sprachmerkmale mit schaltenden Modellen zu berechnen. Zur Modellierung des Rauschens wird ein neues Zustandsmodell eingeführt, für das Expectation-Maximization-(EM-) Algorithmen zur Parameterschätzung hergeleitet werden. Im Back-End werden die statistischen Abhängigkeiten zwischen den Sprachmerkmalen auf der Ebene der Mischungskomponenten des HMMs approximiert. Für das resultierende segmentelle HMM wird eine effiziente Suchstrategie entwickelt. Ein weiterer Schwerpunkt der Untersuchungen ist der Austausch von Informationen zwischen der Stufe der Merkmalsentstörung und der Erkennung. Die Unsicherheit der extrahierten Sprachmerkmale kann bei der Erkennung durch ein sogenanntes Uncertainty Decoding berücksichtigt werden. Daneben ist es möglich, die Informationen aus dem Erkennung in einem mehrstufigen Erkennungsansatz bei der Merkmalsentstörung der jeweils nächsten Erkennungsstufe auszunutzen, wozu die Verteilung der Sprachmerkmale mit Informationen über die HMM-Zustände beeinflusst werden kann. In diesem Ansatz können zur Berücksichtigung der Unsicherheit der Erkennungsergebnisse anstelle einer einzelnen Zustandsfolge, die bei der Standarderkennung berechnet wird, die a posteriori Wahrscheinlichkeiten der HMM-Zustände ermittelt werden. Zu diesem Zweck wird eine wortgraphbasierte Erkennungsstruktur entwickelt, in der zunächst für jeden Sprachrahmen die wahrscheinlichsten Wörter mit einem Vorwärts-Rückwärts-Algorithmus auf Wortebene bestimmt werden. Die a posteriori Wahrscheinlichkeiten der HMM-Zustände können auf einem eingeschränkten Wortgraphen mit einem Vorwärts-Rückwärts-Algorithmus auf Zustandsebene berechnet werden.