

# Noise-robust HMM-based Pattern Recognition using Multimodal Features and Observation Uncertainties

Ahmed Serag Eldin Hussen Abdelaziz

Fortschritte in der Entwicklung von Hidden-Markov-Modell (HMM)-basierter automatischer Mustererkennung haben eine breite Nutzung automatischer Maschinenlernsysteme in vielen modernen Anwendungen ermöglicht. Zu diesen Anwendungen zählen automatische Spracherkennung und automatische Fehlerdiagnose.

In Umgebungen mit akustischen Störungen leiden Systeme wie automatische Spracherkennung stark unter Eingangssignalverzerrungen, die ihre Leistung dramatisch reduziert. Der Schwerpunkt dieser Arbeit ist es, anwendungsspezifische Ansätze zur Erhöhung der Robustheit der automatischen Mustererkennungssysteme gegenüber verschiedenen Rauschtypen zu entwickeln.

Im Bereich der automatischen Spracherkennung (ASR) sind eine Reihe von sogenannten beobachtungsunsicherheitsbasierten Decodierungsansätze vorgeschlagen worden, um die notwendige Robustheit zu erzielen. Die Grundidee dieser Ansätze ist, dass die Fehlanpassung zwischen den gestörten akustischen Beobachtungen und dem zugrundeliegenden statistischen Modell dadurch kompensiert wird, dass die beobachteten verzerrten akustischen Daten nicht als deterministisch sondern als Zufallsvariablen mit einem zeitlich variierenden Unsicherheitsgrad betrachtet werden. Je größer die Unsicherheit eines bestimmten akustischen Merkmals ist, desto kleiner wird ihr Beitrag zur Erkennung.

Eine andere Strategie zur Immunisierung der ASR-Leistung gegenüber Hintergrundrauschen ist es, zusätzlich zu den akustischen Merkmalen eine andere sprachrelevante Informationsquelle zu verwenden, die unabhängig von akustischen Störungen ist. In dieser Arbeit wurde das Videosignal der Lippenbewegungen des Sprechers als diese Informationsquelle betrachtet.

In der Dissertation wurden neue Ansätze, die auf Beobachtungsunsicherheiten und multimodalen Merkmalen basieren, vorgeschlagen, um die Robustheit der HMM-basierten Mustererkennungssysteme gegenüber Hintergrundrauschen zu erhöhen. Mit diesen Ansätzen wurden signifikante Leistungsverbesserungen in Anwendungen wie der automatischen Spracherkennung, Sprachverbesserung und der automatischen Fehlerdiagnose erzielt.

Eine mögliche Erweiterung dieser Studie wäre es, die vorgeschlagenen Ansätze in anderen Anwendungen wie der maschinellen Sprachübersetzung und der Bayes-Netzwerk-basierten Sprecherlokalisierung zu nutzen. Eine weitere Forschungsrichtung ist die breitere Nutzung von neuronalen Netzwerken in den verschiedenen Teilen der vorgeschlagenen Ansätze.