



Zusammenfassung der Dissertation:

**Modellbasierte Merkmalsverbesserung  
zur robusten automatischen Spracherkennung  
in Gegenwart von Nachhall und Hintergrundstörungen**

**des Herrn Alexander Krüger**

Bei der Verwendung von Freisprechsystemen zur automatischen Spracherkennung tritt das Problem auf, dass neben dem gewünschten Sprachsignal sowohl dessen Nachhall als auch unerwünschte Hintergrundstörungen durch das Mikrofon aufgenommen werden. Dieses bewirkt eine Veränderung der für die Dekodierung relevanten akustischen Merkmale, welche aus dem Mikrofonsignal extrahiert werden. Da für das Training des Spracherkenners gewöhnlich ungestörte Sprachsignale herangezogen werden, führt die Diskrepanz zwischen den Trainings- und Einsatzbedingungen zu einer erhöhten Wortfehlerrate.

In dieser Arbeit wird ein neuartiges, auf der Anwendung von Bayes'scher Inferenz basierendes Verfahren zur Verbesserung akustischer Merkmale zur robusten automatischen Spracherkennung in Gegenwart von Nachhall sowie Hintergrundstörungen entwickelt.

Das Verfahren nutzt einerseits a priori-Modelle zur Beschreibung der Trajektorien der akustischen Merkmale jeweils des sauberen Sprachsignals sowie des Hintergrundstörsignals, wobei im ersteren Fall schaltende, lineare dynamische Modelle zur Ausnutzung von Korrelationen zwischen zeitlich aufeinander folgenden Merkmalen eingesetzt werden. Die Arbeit konzentriert sich dabei auf das Training der Modelle sowie die Initialisierung deren Parameter.

Andererseits wird für die Merkmalsverbesserung ein Beobachtungsmodell verwendet, welches die Merkmale des verhallten und gestörten Sprachsignals mit denen des sauberen in Beziehung setzt. Diese Beziehung hängt von der Raumimpulsantwort zwischen dem Sprecher und dem Mikrofon ab. Da deren blinde Schätzung, die in einer unbekanntem Umgebung erforderlich ist, hoch sensibel ist, wird die Raumimpulsantwort durch ein statistisches Modell beschrieben. Dieses besitzt nur zwei Parameter, welche einfacher und robuster als die vollständige Raumimpulsantwort geschätzt werden können.