Forschungsberichte aus dem Arbeitsbereich Nachrichtentechnik
der Universität Bremen

Band 16

Volker Mildner

Signalverarbeitungskonzepte zur robusten Sprechererkennung

D 46 (Diss. Universität Bremen)

Shaker Verlag
Aachen 2007

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Zugl.: Bremen, Univ., Diss., 2007

Printed in Germany.

ISBN 978-3-8322-6504-5
ISSN 1437-000X

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen
Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9
Internet: www.shaker.de • E-Mail: info@shaker.de

Zusammenfassung der Dissertation

Signalverarbeitungskonzepte zur Robusten Sprechererkennung

Die Arbeit beschäftigt sich mit der Aufgabenstellung, ob und zu welchem Grad eine Sprechererkennung unter dem Einfluss akustischer Störungen (Hall, Störgeräusche) möglich ist. Entscheidend ist die Frage, welche Algorithmen der digitalen Signalverarbeitung geeignet sind, um eine Kompensation der Störungen zum Ziel der Erhöhung der Erkennerrate eines Sprechererkennungssystems durchzuführen.

Um die Zusammenhänge zwischen akustischen Störungen, anschließender digitaler Signalverarbeitung und einem darauf folgenden System zur Sprechererkennung detailliert betrachten zu können, werden zunächst die Grundprinzipien eines Systems zur Sprechererkennung erläutert. Hierbei wird auf die einzelnen Teilsysteme der Merkmalsextraktion, der Modellberechnung und der Klassifikation eingegangen. Die zunächst verwendeten Merkmale zur Repräsentation der charakteristischen akustischen Eigenschaften eines Sprechers sind die *Mel Frequency Cepstral Coefficients* (MFCC).

Anhand von Erkennerraten wird die Leistungsfähigkeit des verwendeten Sprechererkennungssystems aufgezeigt. Ferner werden modifizierte Merkmale untersucht, deren Verwendung im ungestörten Fall als eine Steigerung der Erkennerraten ermöglicht. Darauffolgend wird untersucht, welchen Einfluss zusätzliche Störgeräusche auf die Merkmale und somit auf die Erkennerrate haben. Entgegen den Erwartungen erweisen sich gerade jene Merkmale als besonders sensitiv gegenüber Störungen, welche im ungestörten Testfall auf die höchsten Erkennerraten führen. Daher muss untersucht werden, inwiefern sowohl einkanalige als auch mehrkanalige Verfahren der Signalverarbeitung den Einfluss von Störgeräuschen kompensieren können.

Aus dem Bereich der einkanaligen Signalverarbeitung werden die Standardverfahren betrachtet. Die Untersuchungen zeigen, dass einkanalige Verfahren nur sehr bedingt als Vorverarbeitung zur Signalaufbereitung im Zusammenhang mit einem System zur Sprechererkennung anwendbar sind. Entscheidend ist die durch die einkanaligen Verfahren verursachte Verzerrung des Nutzsignals, welche sich nachteilig auf die Merkmale und somit auf die Erkennerrate auswirkt.

In einem weiteren Abschnitt wird die Anwendbarkeit mehrkanaliger Verfahren zur Störgeräuschreduktion betrachtet. Die einzelnen Verfahren machen hierbei unterschiedliche Annahmen bezüglich der Eigenschaften des an den Mikrofonen anliegenden Geräuschfeldes. Alle mehrkanaligen Verfahren haben gemein, dass ihre Filterfunktionen nicht nur hinsichtlich einer Verbesserung des Signal-zu-Rausch Verhältnisses entworfen werden, sondern dass bei dem jeweiligen Entwurf die Nebenbedingung der unverzerrten Übertragung eines Nutzsignals eingehalten wird. Daher sind die mehrkanaligen Verfahren in der generell in der Lage, eine Erhöhung der Erkennerrate zu bewirken, weshalb sie eindeutig den einkanaligen Verfahren vorzuziehen sind.

Im letzten Abschnitt der Arbeit werden neue Merkmale zur Sprechererkennung vorgestellt, welche auf spektro-temporalen Mustern beruhen, die als *Gaborfilter* bezeichnet werden. Derartige Muster sind in der Lage, Modulationen eines Sprachsignals sowohl in Zeit- als auch in Frequenzrichtung zu modellieren, und werden durch ihre jeweiligen Parameter definiert. Durch Analyse des Spektrogramms eines Sprachsignals anhand von Gaborfiltern lassen sich neue Merkmale gewinnen. Um die Parameter der Gaborfilter für die Aufgabe der Sprechererkennung anzupassen, wird eine neue Optimierungsregel vorgestellt. Anschließend werden Gabormerkmale extrahiert und zur Sprechererkennung verwendet. Es zeigt sich, dass bereits ohne Anwendung zusätzlicher digitaler Signalverarbeitung die Gabormerkmale im ungestörten als auch im gestörten Fall höhere Erkennerraten ergeben, als sie durch die leistungsfähigsten MFCC Merkmale erzielt werden können. Eine signifikante Erhöhung der Erkennerrate ist möglich, wenn zusätzlich eine Vorverarbeitung durch mehrkanalige Systeme durchgeführt wird.

Im Rahmen der in dieser Arbeit durchgeführten Untersuchungen zeigt sich, dass durch die gemeinsame Verwendung von mehrkanaligen Verfahren zur Störgeräuschreduktion und anschließender Extraktion von Gabormerkmalen signifikante Gewinne in Bezug auf die Robustheit eines Sprechererkennungssystems erzielt werden können.