

## 3D-Gesichtsbewegungen beim Sprechen

Die Erkennung von Sprache und gesprochenem Inhalt findet heutzutage oft Anwendung in verschiedenen Bereichen wie beispielsweise in digitalen Sprachassistenten. Dabei wird neben der Erkennung des gesprochenen Inhalts, auch der Sprecher anhand akustischer Merkmale identifiziert. Viele solcher Systeme funktionieren jedoch nur auf der akustischen Domäne. Tonaufnahmen können wegen Umgebungsgeräuschen an Qualität verlieren und deswegen Nachteile mit sich bringen. Das Einbeziehen von visuellen Merkmalen als Unterstützung zum akustischen Signal findet deswegen viel Interesse in der Forschung und Entwicklung.

Im Rahmen dieser Arbeit wurde ein System entwickelt, um Gesichtsbewegungen zu analysieren und Zusammenhänge zwischen Gesichtsbewegungen beim Sprechen und akustischen Merkmalen zu untersuchen. Hierzu wurden Gesichtsbewegungen mittels strukturierter Lichts und einer Hochgeschwindigkeitskamera aufgezeichnet und Audio Metriken wie MFCC, Mel-Leistung, Pitch und Peak Hüllkurve des Audiosignals mit den Bewegungen der projizierten Lichtstreifen sowie deren Breiten im Gesicht korreliert. Die Resultate zeigen, dass MFCC 2 und MFCC 4 über alle Testpersonen die zuverlässigsten Metriken sind, mit Korrelationskoeffizienten im Bereich von 0.6 - 0.85. Darüber hinaus wurden die auftretenden Korrelationen im Gesicht lokalisiert. Das in der vorliegenden Arbeit entwickelte System erlaubt eine Analyse der Gesichtsbewegungen mit einer Bildrate von 300 Bildern pro Sekunde, im Vergleich zu den heutzutage typischen Analysen bei 30 Bildern pro Sekunde. Die erhöhte Wiederholungsrate bietet eine bessere örtliche Auflösung, weil ca. 50 % mehr Lichtstreifenpositionen über lange Zeit damit verfolgt werden können. Des Weiteren führt die erhöhte Wiederholungsrate zu ca. 20 % höheren Korrelationswerten, was die Analyse von Gesichtsbewegungen, welche aufgrund von Artikulation beim Sprechen entstehen, erleichtert. Ein zusätzlicher Vorteil des entwickelten Systems besteht darin, dass keine Marker an das Gesicht geklebt werden müssen. Die zukünftig möglichen Anwendungen des entwickelten Systems umfassen Bereiche wie Biometrie, Sprecheridentifizierung, Phonetik und Forensik.



Diplomierende

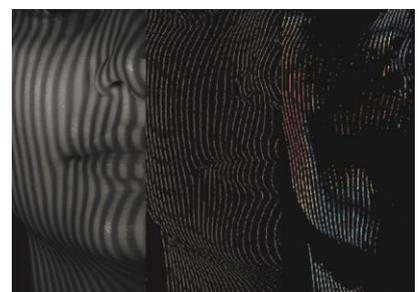
Marjan Laski  
Lucas Oehms

Dozent

Julien Rey



Eine LED Optik projiziert strukturiertes Licht auf das Gesicht einer Person. Mit einer Hochgeschwindigkeitskamera und einem Mikrofon wurden Aufnahmen einer sprechenden Person gemacht, welche nachfolgend in MATLAB bearbeitet und analysiert wurden.



Links ist die Graustufenaufnahme zu sehen, welche im mittleren Bild kantendetektiert wurde. Die Korrelationskoeffizienten aus der Analyse werden im rechten Bild dargestellt, wobei Rot für eine hohe und Blau für eine niedrige Korrelation steht.