

Ob Siri oder Amazon Echo – die automatische Spracherkennung hält Einzug in unser Leben und ist trotz erstaunlicher Verbesserung bei der Erkennung im Allgemeinen noch lange nicht so gut wie das menschliche Sprachverstehen. Um mögliche Wege für robustere und ggfs. verteilte Spracherkennungssysteme zu erschließen, beschäftigt sich der Journal-Bericht "Turbo Automatic Speech Recognition" mit einem neuartigen Verfahren zur iterativen optimalen Informationsfusion. Eine Fusion ist immer dann nötig und gewinnbringend, wenn man unterschiedliche Informationsquellen statistisch optimal kombinieren möchte. Dies kann die Verknüpfung von Audio (Spracherkennung) und Video (Lippenlesen) sein, aber auch die Verknüpfung zweier gleichartiger Sensoren (zwei Mikrofone, oder beim Menschen das rechte und das linke Ohr). Bei dem gewählten Ansatz handelt es sich um die konsequente Anwendung des aus der Nachrichtentechnik berühmten Prinzips der Turbo-Codes auf Fragen der automatischen Spracherkennung mit mehreren Datenströmen. Als Hauptneuerung wird in der Publikation ein sogenannter modifizierter Viterbi-Algorithmus vorgestellt, der eine neuartige Informationsrepräsentation zur iterativen Rückkopplung vorsieht. Zwei Einzel-Erkenner werten dabei wiederholt ihr jeweiliges Eingangssignal der zugrunde liegenden Sprachäußerung aus und tauschen von Iteration zu Iteration Informationen miteinander aus, und hangeln sich damit Schritt für Schritt hin zu einem gemeinschaftlich verbesserten Erkennungsergebnis.