

Spracherkennung liest von den Lippen ab

Forscher der Carnegie Mellon University (<http://www.cmu.edu>) in Pittsburgh (USA) haben eine Software namens NLips entwickelt, die visuelle und akustische Informationen zur Spracherkennung nutzt. Wie herkömmliche Programme teilt NLips gesprochene Passagen in einzelne Abschnitte, so genannte Phoneme. Simultan registriert eine Kamera die Lippenbewegungen, die ein neuronales Netzwerk erkennt und verarbeitet. Das optische Äquivalent zu den Phonemen sind Viseme, also Teilabschnitte der Bewegungsabfolge beim Spre-

chen. 50 solcher Viseme dienen als Grundlage für diese Art der Spracherkennung. Ein Vorteil der Software ist, dass sie laut ihrer Entwickler auch in Räumen mit verhältnismäßig hoher Hintergrundlautstärke einsetzbar ist. Hier würden Programme, die nur Phoneme auswerten, deutlich größere Schwierigkeiten haben. Bei ruhiger Umgebung schneidet NLips mit einer Erfolgsrate von 92 Prozent zwar nicht entscheidend besser ab als gängige Programme. Bei starkem Hintergrundlärm sinkt diese Rate jedoch nur auf 85 Prozent

ab, während rein akustisch arbeitende Systeme dort den Aussagen eines Entwicklers zufolge nur rund 60 Prozent erreichen.

Bislang hat man NLips beim Buchstabieren von Einzelwörtern getestet. Die Wissenschaftler sind zuversichtlich, dass auch der nächste Schritt, also das Interpretieren eines kontinuierlichen Sprachflusses, keine Probleme bereiten wird. In Zukunft will man NLips in ein Videokonferenzsystem integrieren, das automatisch Mitschriften von Diskussionen anfertigen soll. (gas)