



**Ruprecht-Karls-Universität Heidelberg**  
**Medizinische Fakultät Mannheim**  
**Dissertations-Kurzfassung**

**Auswertung der Sprachaudiometrie mittels  
Spracherkennungssoftware, basierend auf dem Oldenburger  
Satztest**

Autor: Christian Warken  
Institut / Klinik: Klinik für Hals-Nasen-Ohrenheilkunde, Kopf- und Halschirurgie  
Doktormutter: Prof. Dr. A. Schell

Sprache ist ein wesentliches Element der Kommunikation und Interaktion innerhalb sozialer Gesellschaften. Die Sprachaudiometrie gilt als entscheidende Diagnostik für die quantitative Bemessung eines möglichen Hörschadens. Zur Überprüfung der Sprachverständlichkeit stehen im deutschsprachigen Raum eine große Anzahl verschiedener Testverfahren zur Verfügung, die das Verstehen von Einsilbern, Mehrsilbern oder ganzer Sätze testen.

Hörgeschädigte Personen haben beim Sprachverständnis vor allem in lauter Umgebung weitaus größere Probleme als Normalhörenden. Um diesen Umstand realistisch zu erfassen, wird mit dem Oldenburger Satztest (OLSA) bei schwerhörigen Patienten im Störgeräusch geprüft. Der OLSA ist ein adaptiver Sprachtest, der Pegel der Sprache wird hierbei entsprechend der Antwort der Testperson verändert.

Die automatische Reintonaudiometrie liefert ein genaues Maß der Hörschwelle, aber Validierungsdaten für andere audiometrische Verfahren sind weiterhin begrenzt. Zudem besteht eine große Diskrepanz zwischen dem Bedarf und der Kapazität von Audiologen für die Durchführung von Hörtests. Durch eine Automatisierung könnten Zeitersparnisse entstehen und mehr Patienten von der Diagnostik profitieren. Sprecherabhängige Spracherkennungsprogramme werden vom Benutzer auf Eigenarten der Aussprache trainiert. In Abhängigkeit von der Erfahrung und dem Trainingsumfang können sehr hohe Erkennungsquoten erreicht werden. Die neuste Generation dieser Systeme müssen nicht mehr zwangsläufig trainiert werden. Neben dem Umfang und der Flexibilität des zugrunde liegenden Wortschatzes des Spracherkennungssystems, spielt auch die akustische Qualität des Inputs eine entscheidende Rolle. Die wichtigste Voraussetzung ist hierbei zweifelslos eine präzise, deutliche und flüssige Aussprache.

Wir konnten zeigen, dass die Auswertung des OLSA mit der Spracherkennungssoftware Dragon von Nuance Inc. signifikant schlechter war als die manuelle Auswertung durch geschultes Fachpersonal ( $p < 0,0001$ ). Die Auswertung der gesamten Studienkohorte ergab ein heterogenes Bild, mit Testreihen die deckungsgleich zur Auswertung durch die MTA waren und Testungen mit Abweichungen von über 50%. Limitierend für die statistische Auswertung waren die teilweise zu kleinen Subgruppen. Für Folgestudien sollte sich auf genau definierte Probandengruppen konzentriert werden, um ausreichende Gruppengrößen rekrutieren zu können.

Für die gewünschte Fragestellung ist die Auswahl des geeigneten Spracherkennungssystem essenziell. Das „perfekte“ Spracherkennungsprogramm existiert nicht. Durch eine Trainingsphase kann die Qualität und Verwertbarkeit der Ergebnisse signifikant verbessert werden. In unserer Studie fehlte die Auswertung der größten Vergleichsgruppe, nämlich Menschen ohne Hörschädigung.

In den vergangenen Jahren wurden zunehmend Versuche unternommen, neuronale Netze für das akustische Modell zu verwenden. Moderne Spracherkennungssysteme, die auf Deep Learning aufsetzen, verbessern ihr Outcome kontinuierlich und liefern Erkennungsraten, die in Bereichen liegen, die auch ein gesunder Mensch erreichen würde. In der aktuellen Forschung zu Deep-Learning-Systemen in der Spracherkennung werden verschiedene stochastische Modelle modifiziert und miteinander verknüpft.