

# Ein Modell der Diskriminationsleistung im Störgeräusch bei Cochlea-Implantaten mit elektrisch-akustischer Stimulation (EAS)

*Tobias Rader<sup>1,2</sup>, Youssef Adel<sup>2</sup>, Uwe Baumann<sup>1</sup> und Hugo Fastl<sup>2</sup>*

1) Audiologische Akustik, Klinik für HNO-Heilkunde, Goethe-Universität Frankfurt am Main

2) AG Technische Akustik, MMK, Technische Universität München

## Einleitung

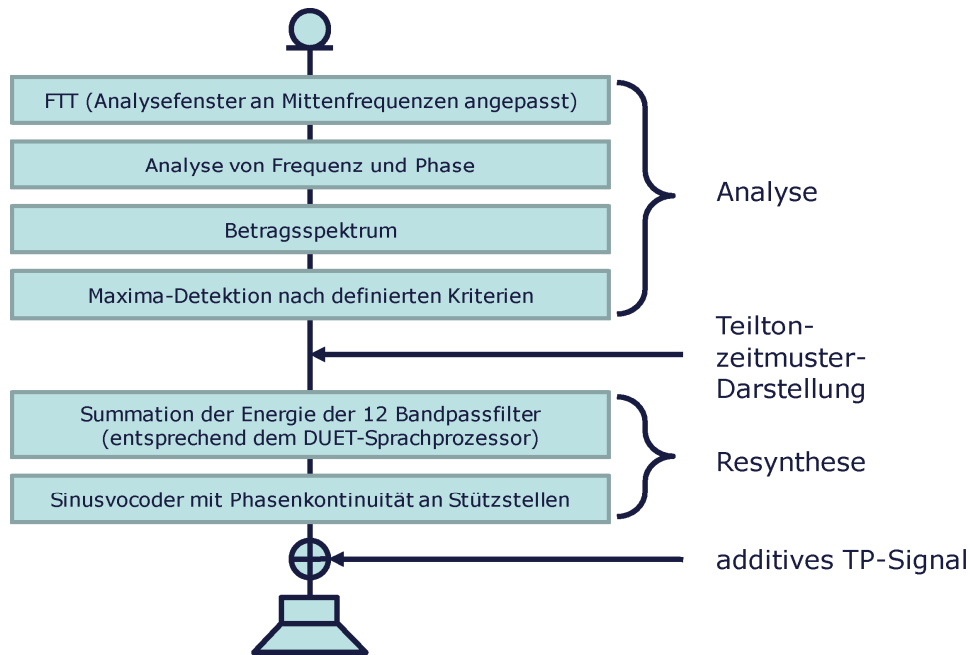
Mit elektrisch-akustischer Stimulation (EAS) bezeichnet man die simultane Erregung des Hörnervs mit einerseits einem Cochlea-Implantat (CI) für Frequenzen oberhalb von etwa 500 Hz und andererseits der Nutzung des niederfrequenten Restgehörs unterhalb von etwa 500 Hz am selben Ohr. Eine wichtige Voraussetzung für den Erfolg einer EAS-Versorgung ist ein stabiles Restgehör in den tiefen Frequenzen. Patienten mit einer EAS-Versorgung zeigen ein signifikant besseres Verstehen im Störgeräusch (Rader et al. 2009) gegenüber konventionell versorgten CI-Patienten. Dieser Effekt wird mit CI/EAS-Simulationen am Modell nachvollzogen und mit den Daten von EAS-versorgten CI-Patienten verglichen.

## Material und Methoden

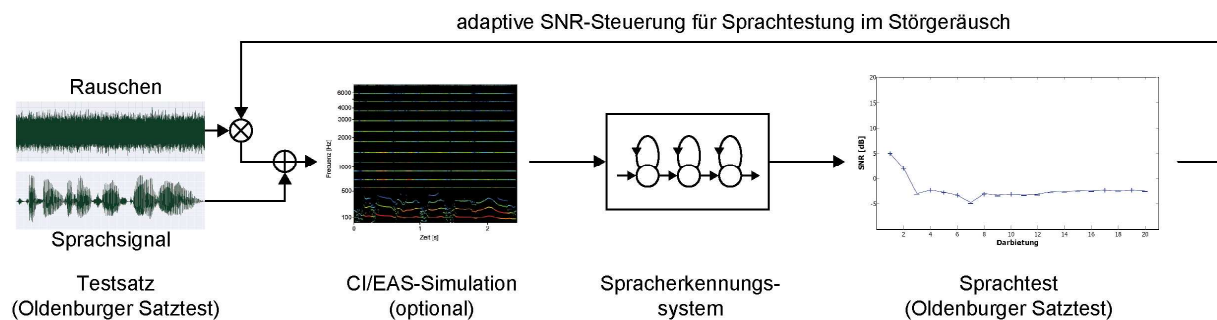
Das Sprachmaterial des Oldenburger Satztests (OLSA, Wagener et al. 1999) wurde spektral transformiert und durch Signalverarbeitung/Resynthese in eine CI/EAS-Simulation übergeführt (Rader et al. 2010). Ein Blockdiagramm der Signalverarbeitung ist in Abbildung 1 dargestellt. Zur Simulation der über die Elektroden übertragenen Sprachanteile wurde das Sprachsignal oberhalb der Übergabefrequenz entsprechend der zwölf Mittenfrequenzen des Med-El DUET Sprachprozessores bandpassgefiltert. Der CI-Übertragungsbereich beträgt 500 Hz bis 8500 Hz. Die Mittenfrequenzen der CI-Bandpasskanäle betragen dabei 567 Hz, 717 Hz, 909 Hz, 1150 Hz, 1457 Hz, 1845 Hz, 2336 Hz, 2958 Hz, 3746 Hz, 4743 Hz, 6007 Hz und 7606 Hz. Die Resynthese des Sprachmaterials erfolgte mit einem 12-Band-Sinusvocoder, bei dem die Vocoder-Frequenzen mit den 12 Mittenfrequenzen der CI-Bandpasskanäle übereinstimmen. Die Störgeräusche (Fastl-Rauschen und OL-noise) wurden gleichermaßen in einer CI/EAS-Simulation prozessiert. Durch Variation der Grenzfrequenz des tieffpassgefilterten Anteils  $f_p$  kann das Signal von einer reinen CI-Simulation ( $f_p=0$  Hz) zu einer EAS-Simulation mit variablen akustischen tieffrequentem Restgehör variiert werden. Die Berechnung der Schalle erfolgte in der Teiltonzeitmysterdomäne nach Heinbach (1988).

Um den Einfluss des Tiefpass (TP)-Anteils eines EAS-Systems auf die Sprachdiskrimination zu untersuchen, wurde ein auf dem Hidden-Markov Modell basierendes, automatisiertes Spracherkennungssystem (HTK-Toolkit, Young et al. 2006) implementiert. Mit Hilfe einer Beschreibungssprache des HTK wurde die Grammatik der OLSA-Sätze beschrieben und das Spracherkennungssystem mit den unverarbeiteten OLSA-Sätzen trainiert. Der Frequenzbereich des Spracherkennungssystems wurde dabei auf eine obere Grenzfrequenz von 8 kHz - entsprechend dem MED-EL DUET-Sprachprozessor - beschränkt.

Als Rauschsignale kamen das quasi unmodulierte OLSA-Rauschen (OL-noise, 30 verschobene und überlagerte Sätze des OLSA) und das sprachmodulierte Fastl-Rauschen (mit stochastisch verteilten Pausen) zum Einsatz. Beide Rauschsignale beinhalten keine verwertbaren Sprachinformationen (d.h. kein „informational masking“).



**Abbildung 1:** Blockdiagramm der Signalverarbeitung zur EAS-Simulation. Die Sinuston-Vocoder Resynthese erfolgte entsprechend der Mittenfrequenzen des Med-El DUET-Sprachprozessors, die Grenzfrequenz des additiven Tiefpass-Signals (TP) wurde von 0 bis 500 Hz variiert.



**Abbildung 2:** Modell zum Sprachverstehen im Störgeräusch. Je nach Erkennungsrate des Spracherkenners wird das Störgeräusch adaptiv angepasst, um die Sprachverständlichkeitsschwelle zu ermitteln.

Die Sprachverständlichkeitsschwelle (SVS), also der SNR-Wert, bei dem 50 % des Satzes richtig erkannt wurde, wurde mit Hilfe des Oldenburger Satztestes in Verbindung mit dem HTK ermittelt. Das Störgeräusch wurde adaptiv entsprechend der aktuellen Erkennungsrate variabel dargeboten, wobei das Sprachsignal im Pegel konstant blieb.

## Ergebnisse

Abbildung 3 links zeigt die Diskriminationsfunktionen des HTK für Fastl-Rauschen und OL-noise mit unbearbeiteten Schalldateien (ohne EAS-Simulation). Die SVS liegt dabei für Fastl-Rauschen bei -7.5 dB SNR mit einer Steigung (approximiert im linearen Teil der Diskriminationsfunktion) von 6.6 % pro dB SNR. Mit dem Störgeräusch OL-noise beträgt der Wert der SVS 3.4 dB SNR und der Steigung 8.4 % pro dB SNR. Der SNR-Abstand zwischen dem kontinuierlichem und dem sprachsimulierenden Rauschen beträgt 10.9 dB SNR.

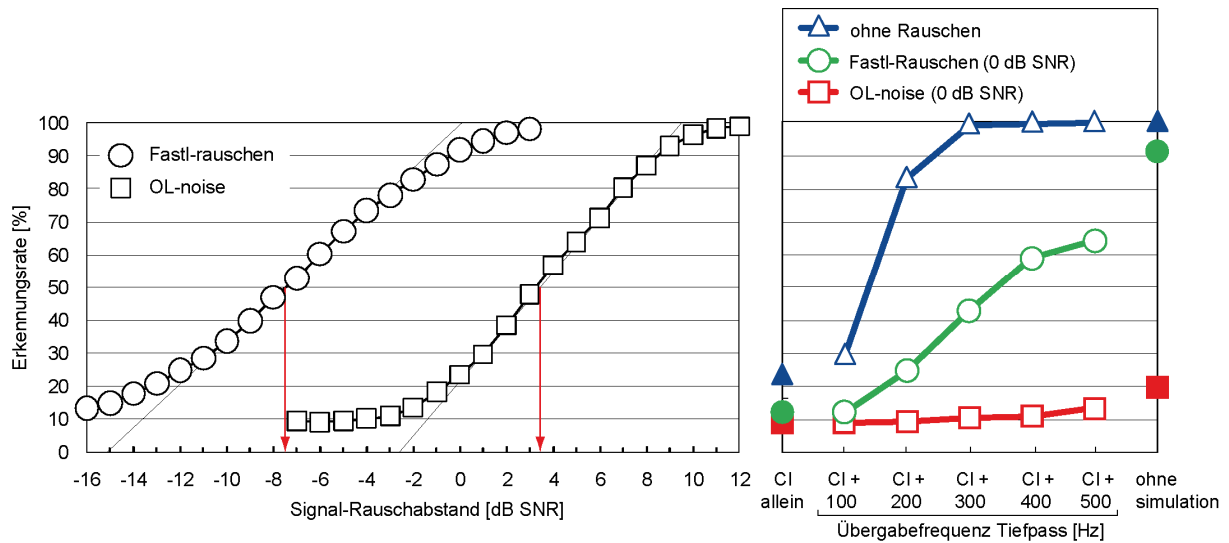


Abbildung 3 links: Diskriminationsfunktionen des Oldenburger Satztests für Fastl-Rauschen (Kreise) und Oldenburger Rauschen (Quadrate) gemessen am Modell (unbearbeitete Schallsignale, ohne CI/EAS-Simulation). Die roten Pfeile zeigen die Sprachverständlichkeitsschwellen (Signal-Rauschabstand für die Erkennungsrate 50 %), die dünnen Linien repräsentieren die Steigung im linearen Teil der Diskriminationsfunktion.

Abbildung 3 rechts: Erkennungsraten des Spracherkenners mit Signalen der CI/EAS-Simulationen in folgenden Konditionen: „in Ruhe“ (Dreiecke), sowie Fastlrauschen (Kreise) und OLSA-Rauschen (Quadrate) bei jeweils 0 dB SNR. Offene Symbole zeigen den Einfluss eines additiven TP-Anteils (EAS-Simulation), gefüllte Symbole die Konditionen „CI allein“ und „ohne Simulation“.

Der Einfluss tieffrequenter Signalanteile auf die Erkennungsrate des Spracherkennungssystems HTKs im Störgeräusch ist in Abbildung 3 rechts dargestellt. Bei Verwendung von bearbeiteten Schallsignalen entsprechend einer CI-Simulation ohne akustischen, tieffrequenten Anteil (Kondition „CI allein“) bricht die Erkennungsrate in den drei getesteten Konditionen „ohne Rauschen“, „Fastlrauschen“ und „OL-noise“ deutlich ein. Die Ergänzung der CI-Simulations-Signale um ein additiv hinzugefügtes TP-gesfiltertes Sprachsignal bewirkt bei der tiefsten Filterfrequenz ( $f_{tp}=100$  Hz) bei beiden Störgeräuscharten noch keinen Effekt, in Ruhe eine leichte Verbesserung der Erkennungsrate um etwa 8 % (verglichen zur Kondition „CI allein“). Erhöht man die TP-Grenzfrequenz auf  $f_{tp}=200$  Hz, zeigt sich eine deutliche Erhöhung der Erkennungsrate um 60 % „ohne Rauschen“ und um 15 % für Fastl-Rauschen. Ab  $f_{tp}=300$  Hz konvergiert die Erkennungsrate „ohne Rauschen“ gegen 100 %, wobei sich für das Störsignal Fastl-Rauschen ein weiterer linearer Anstieg der Erkennungsrate bis 400 Hz ergibt, der erst bei einer akustischen Übergabefrequenz des Tiefpasses  $f_{tp}=500$  Hz leicht konvergiert.

## Zusammenfassung

Das Modell zeigt einen Ansatz zur Abschätzung des möglichen Sprachverstehens in Abhängigkeit vom nutzbaren tieffrequenten Restgehörs bei Patienten mit einer EAS-Versorgung. Die mit dem Spracherkennersmodell gewonnenen Ergebnisse verdeutlichen den größten Nutzen tieffrequenter akustischer Stimulation bei sprachmoduliertem Störgeräusch ab einer Tiefpassfilterfrequenz oberhalb von 300 Hz. Erwartungsgemäß ermöglichen die Pausen im sprachmodulierten Rauschen nach Fastl auch dem Spracherkennungssystem eine günstigere Sprachdiskrimination im Vergleich zu einem kontinuierlichen Sprachsignal.

## Danksagung

Die Autoren danken Felix Weninger und Martin Wöllmer vom Lehrstuhl für Mensch-Maschine-Kommunikation der TU-München für die Bereitstellung der Sprachkorpora und die Unterstützung bei der Adaption des HTK-Toolkit.

## Referenzen

- Rader T, Baumann U and Fastl H. Comparison of Speech Intelligibility by EAS, Bimodal, Uni- and Bilateral Cochlear Implant Patients in a 'Multi-Source Noise Field' (MSNF). In: M. Boone, editor, Proc. Intern. Conf. on Acoustics (NAG/DAGA 2009), Rotterdam, The Netherlands, pp. 1288–1290, 2009.
- Rader T, Baumann U, und Fastl H.: Simulationen zum Sprachverstehen im Störgeräusch bei elektrisch-akustischer Stimulation (EAS). In: Tagungsband der 13. Jahrestagung der Deutschen Gesellschaft für Audiologie, Frankfurt. DGA, 2010.

- Heinbach W. Aurally adequate signal representation: The Part-Tone-Time-Pattern. *Acustica* 67, 113-121, 1988.
- S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland: *The HTK Book (for HTK version 3.4)*, Cambridge University Engineering Department, 2006.
- Wagner K, Brand T, Kuehnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests in deutscher Sprache I – III: Design, Optimierung und Evaluation des Oldenburger Satztests . *Z Audiol*; 38 (1-3) : 4 – 15, 44 – 56, 86 – 95, 1999.