

# Kombination einkanaliger und mehrkanaliger Algorithmen zur Störgeräuschreduktion

Thomas Rohdenburg, Volker Hohmann

Medizinische Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg

Email: thomas.rohdenburg@uni-oldenburg.de

## Einleitung

In dieser Studie werden einkanalige state-of-the-art Algorithmen zur Störgeräuschreduktion mit Hilfe des perceptiven Qualitätsmaßes PEMO-Q nach Huber [5] bewertet und optimiert. Dieses basiert auf dem psychoakustischen Perzeptionsmodell (PEMO) nach Dau et al. [2]. Die hier betrachteten Verfahren gehören zur Klasse der STSA<sup>1</sup>-Algorithmen und versuchen die Einhüllende des Nutzsignals in Sub-Bändern mit Hilfe eines zeitvarianten Filters im Frequenzbereich zu rekonstruieren. Die Algorithmen enthalten Parameter, mit denen die Stärke der Störgeräuschreduktion kontrolliert werden kann. Eine Maximierung der Störgeräuschdämpfung hat jedoch im Allgemeinen eine Verzerrung des Nutzsignals zufolge, die nur in gewissem Ausmaß toleriert werden kann. Das häufig zur Bewertung der Störgeräuschreduktion verwendete Signal-to-Noise-Ratio Enhancement (SNRE) kann keine Aussage über die Verzerrung des Nutzsignals machen. Ebenfalls als ungeeignet erweisen sich lineare Bewertungsmaße, wie die Korrelation zwischen dem geschätzten und dem wahren Nutzsignal. Für die Qualitätsbewertung besser geeignet ist das perceptiv Ähnlichkeitsmaß PSM<sup>2</sup> aus PEMO-Q, welches sich durch eine hohe Übereinstimmung zwischen der subjektiv wahrgenommenen und der gemessenen Qualität im Bezug auf die Signalverzerrung in Audio-CODECs auszeichnet [5]. Eine objektive Qualitätsbewertung der STSA-Algorithmen zeigte, dass die Algorithmen nicht robust gegen instationäre und spektral-gefärbte Störungen, wie z.B. Cafeteria-Rauschen sind. Diese Erkenntnis legt die Verwendung mehrkanaliger Störgeräuschreduktionsverfahren nahe. Im Rahmen dieser Arbeit wurden daher mehrkanalige Verfahren unter Einbeziehung des perceptiven Qualitätsmaßes verglichen.

## Einkanalige Algorithmen

Als einkanalige state-of-the-art Algorithmen wurden die STSA-Algorithmen nach Ephraim und Malah [4, 3] verwendet. Diese zeichnen sich durch eine hohe Störgeräuschreduktion aus und verringern die bei spektralen Subtraktionsverfahren als *Musical Tones* bekannte Reststörung. Eine anschauliche Interpretation der beteiligten Filter-Parameter findet sich in Cappé [1]. Die wichtigsten Parameter sind zwei SNR-Schätzwerte: Ein instantan gemessenes (a posteriori) Signal-zu-

Rauschverhältnis, sowie ein a-priori SNR-Schätzwert, der durch eine rekursive Glättung vorangegangener a posteriori-Werte berechnet wird.

Die betrachteten Algorithmen benötigen eine zuverlässige Schätzung der Störgeräuschleistung. Hierfür wird das Minimum-Statistics-Verfahren nach Martin [6] verwendet.

Der rekursive Glättungsparameter des a-priori SNR-Schätzwertes ist ein wichtiger Parameter, der die Stärke der Störgeräuschreduktion beeinflusst<sup>3</sup>. Wird dieser Parameter variiert, kann man unter der Berücksichtigung der Signalqualität eine optimale Glättungszeitkonstante  $\tau$  ermitteln.

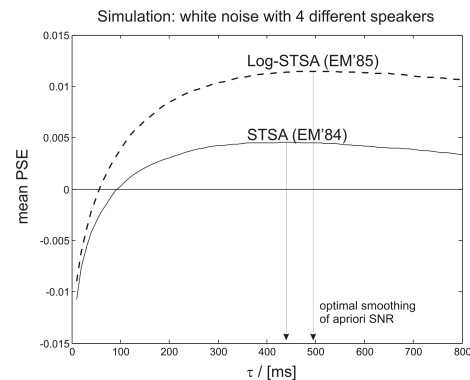


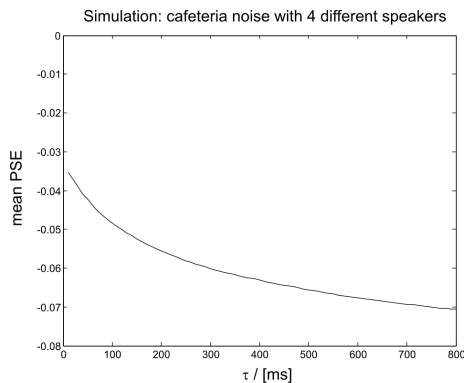
Abbildung 1: Perceptual Similarity Enhancement (PSE) für stationäre Störungen

In Abbildung 1 wurde das PSM-Maß des gefilterten Signals auf das PSM des unverarbeiteten Signals normiert. Dieses neue Maß wird mit Perceptual Similarity Enhancement (PSE) bezeichnet. Für positive PSE-Werte sagt das Modell eine Signalverbesserung gegenüber dem ungefilterten Referenz-Signal voraus. In diesem Bereich ist auch eine Verbesserung der Sprachverständlichkeit zu erwarten. Negative Werte bedeuten hingegen, dass eine Filterung des Signals mit den gegebenen Parametern nicht sinnvoll ist, da sie eine Verringerung der Signalqualität zur Folge hat. Abbildung 1 zeigt, dass für weißes Rauschen als Störgeräusch und Sprache als Nutzsinal die optimale Glättungszeitkonstante im Bereich  $450 \text{ ms} < \tau < 500 \text{ ms}$  liegt und damit in etwa mit den in der Literatur [1] genannten Werten übereinstimmt. Abbildung 2 zeigt die Ergebnisse für Cafeteria-Rauschen, bei denen keine Verbesserung gegenüber der unverarbei-

<sup>1</sup>Short Time Spectral Attenuation

<sup>2</sup>Perceptual Similarity Measurement

<sup>3</sup>In der Literatur wird dieser Parameter häufig auf einen festen, empirisch ermittelten Wert gesetzt



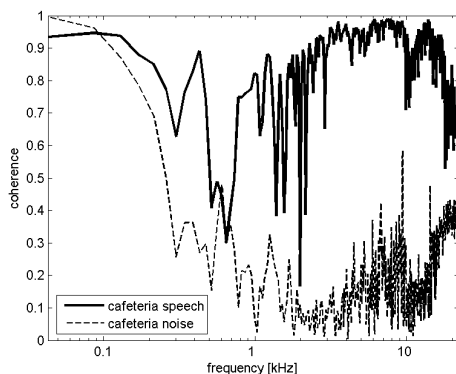
**Abbildung 2:** Perceptual Similarity Enhancement (PSE) für instationäre Störungen (Cafeteria Rauschen)

teten Referenz zu erzielen ist. Die einkanaligen Algorithmen sind somit nicht robust gegenüber instationären, spektral-gefärbten Störgeräuschen.

### Mehrkanalige Algorithmen

Die betrachteten mehrkanaligen Algorithmen basieren auf der Beobachtung, dass zwischen linkem und rechtem Ohr unkorrelierte Signale im allgemeinen als Störgeräusch wahrgenommen werden und daher zu unterdrücken sind [7].

Aus Abbildung 3 ist ersichtlich, dass die Unkorreliertheit zwischen linkem und rechtem Kanal für reale Störgeräusche nur in höheren Frequenzbändern ab 800 Hz zu beobachten ist.



**Abbildung 3:** Kohärenz von Sprache und Störgeräusch zwischen linkem und rechtem Hörgerätemikrofon

Trotz dieser Einschränkungen lassen sich mit diesen Verfahren in der Praxis recht gute Störgeräuschdämpfungen erzielen [7].

Als Kohärenzmaß zwischen dem linken und dem rechten Signal dient eine ERB<sup>4</sup>-basierte Magnitude Squared Coherence (MSC), die aus ERB-Leistungsdichtespektren berechnet werden kann und Werte zwischen 0 und 1 annimmt. Werte nahe 1 sagen eine hohe Korrelation der Kanäle im entsprechenden Frequenzband voraus - und damit einen hohen Nutzsignalanteil, während Frequenzbänder mit einer geringen MSC als rauschhaft betrachtet werden können.

<sup>4</sup>equivalent rectangular bandwidth

Mit diesen Informationen können nun zeitvariante Filter gesteuert werden, die eine Dämpfung des Störsignals bewirken.

1. Beim Binauralen Enthaller-Algorithmus nach Wittkop et al. [7] wird die MSC direkt zur Gewichtung der Frequenzbänder verwendet.
2. Bei einem neuen Verfahren wurde die MSC zur Schätzung des (mehrkanaligen) Störgeräusches verwendet, welches dann mit dem STSA-Algorithmus nach Ephraim und Malah gefiltert wurde.
3. In einem dritten Verfahren wurde der mehrkanalige Störgeräuschschätzer aus (2) mit dem Minimum-Statistics-Verfahren kombiniert.

### Ergebnisse und Zusammenfassung

Mit einem objektiven Qualitätsmaß konnte gezeigt werden, dass einkanalige Störgeräuschreduktionsverfahren offenbar nicht robust gegenüber instationären Störungen sind und in diesen Fällen immer zu einer Verringerung der Signalqualität gegenüber dem unverarbeiteten Referenzsignal führen.

Mehrkanalige SNR-Schätzer können bei hohen Frequenzen eine Störgeräuschreduktion und eine Steigerung der Qualität erzielen.

Das kombinierte Verfahren (3) konnte im Vergleich zu dem bekannten Verfahren (1) eine höhere Störgeräuschdämpfung erzielen, führte aber gegenüber (1) zu einer leicht stärkeren Signalverzerrung (siehe Tabelle 1). Weitere Untersuchungen sollen geeignete Parametersätze entwickeln.

Algo.	SNRE [dB] Links	SNRE [dB] Rechts	PSE Links	PSE Rechts
(1)	2.09	1.92	0.02	0.02
(2)	3.41	2.49	-0.05	-0.07
(3)	4.71	3.29	-0.05	-0.07

**Tabelle 1:** Ergebnisse mehrkanaliger Algorithmen

### Literatur

- [1] O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. on Speech and Audio Processing*, 2(2):345–349, 1994.
- [2] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the ‘effective’ signal processing in the auditory system i. *J. Acoust. Soc. Am*, 99(6), 1996.
- [3] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 32(6):1109–1121, 1984.
- [4] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 33(2):443–445, 1985.
- [5] R. Huber. *Objective assessment of audio quality using an auditory processing model*. PhD thesis, University of Oldenburg, 2003.
- [6] R. Martin. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Trans. on Speech and Audio Processing*, 9(5):504–512, 2001.
- [7] T. Wittkop and V. Hohmann. Strategy-selective noise reduction for binaural digital hearing aids. *Speech Commun.*, 39(1-2):111–138, 2003.