

# Optimierung eines einkanaligen Verfahrens zur Störgeräuschunterdrückung für Kommunikationsanwendungen

Wilhelm Grünbaum, Christian Gruber

voice INTER connect GmbH, Dresden, Email: gruenbaum@voiceinterconnect.de

## Einleitung

Die Reduktion von Störgeräuschen ist für Kommunikationsanwendungen von großer Bedeutung. Viele Kommunikationssysteme wie Freisprecheinrichtungen im KFZ oder Türsprechanlagen sind einkanalig aufgebaut, d.h. dass nur ein Mikrofonsignal zur Verarbeitung zur Verfügung steht. Störsignale werden in diesem Fall ebenso wie Nutzsignale akustisch detektiert und stören die Kommunikation. Das Ziel der Störgeräuschunterdrückung besteht in der Verbesserung der subjektiven Sprachsignalqualität besonders bei geringem Signal-Rausch-Verhältnis, wofür eine zuverlässige Schätzung der Störanteile erforderlich ist. Im vorliegenden Beitrag werden die wesentlichen Probleme einkanaliger Unterdrückungsalgorithmen beschrieben und ein Lösungsansatz vorgestellt, bei dem besonders auf eine möglichst ressourcensparende Umsetzung Wert gelegt wurde.

## Einkanalige Störgeräuschunterdrückung

In einkanaligen Systemen zur Störgeräuschunterdrückung steht nur ein Eingangssignal zur Verarbeitung zur Verfügung. Sämtliche für die Störgeräuschunterdrückung wichtigen Informationen müssen allein aus diesem Signal gewonnen werden. Aus diesem Grund sind einkanalige Verfahren bei weitem nicht so leistungsstark wie mehrkanalige, denen weitere abhängige Signale zur Bestimmung der unabhängigen Störgeräuschkomponente zur Verfügung stehen. Besonders bei instationären Störungen und sehr geringem Signal-Rausch-Verhältnis (SNR) sind einkanaligen Verfahren deutliche Grenzen gesetzt. Den prinzipiellen Aufbau eines einkanaligen Systems zur Störgeräuschunterdrückung zeigt Abbildung 1. Das Eingangssignal besteht aus dem mit ei-

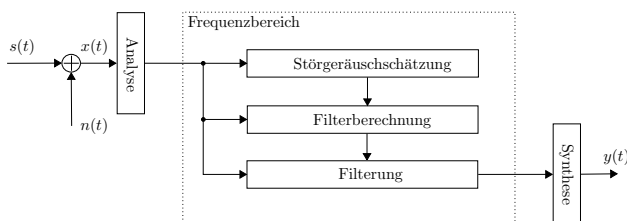


Abbildung 1: Prinzipieller Aufbau einkanaliger Systeme zur Störgeräuschunterdrückung

nem Störsignal überlagerten Sprachsignal. Dieses wird üblicherweise mit der schnellen Fourier-Transformation (FFT) in den Frequenzbereich transformiert. Dort erfolgt die Schätzung des Stör-Leistungsdichtespektrums (LDS). Aus diesem und dem Eingangs-LDS wird die frequenzdiskrete Übertragungsfunktion des Unterdrückungsfilters

$H(\lambda)$  berechnet. Die Filterung erfolgt als Multiplikation im Frequenzbereich. Das Faltungsergebnis wird nun wieder in den Zeitbereich rücktransformiert. Da die FFT für periodische Zeitsignale definiert ist, erfolgt eine Zerlegung des Eingangssignals in Blöcke, die in gedachter periodischer Fortsetzung nacheinander verarbeitet werden. Das Ausgangssignal wird auf ähnliche Weise aus den Ergebnisblöcken zusammengesetzt, wofür das Overlap-Add-Verfahren verwendet werden kann.

## Probleme

Häufig sind bei der Störgeräuschunterdrückung die folgenden Probleme zu beobachten, die hauptsächlich durch mangelnde Kenntnis der Eigenschaften des Störgeräuschs verursacht werden.

- Verzerrung des Sprachsignals auch bei geringem Störpegel
- Artefakte (Musical Tones) im Ausgangssignal besonders bei varianten Störsignalen
- Lange Adaptionszeiten bei sprunghaft einsetzenden und abrupt endenden Störsignalen

## Störgeräuschschätzung

Um die oben genannten Effekte möglichst gering zu halten, sollte das Schätzverfahren die folgenden Anforderungen erfüllen:

- Keine Unterschätzung des Stör-LDS in reinen Störabschnitten
- Keine Überschätzung des Stör-LDS in Sprachabschnitten
- Möglichst geringe Adaptionszeiten

Zudem sollte die Schätzung des diskreten Stör-Leistungsdichtespektrums (LDS)  $\Phi_{nn}(\lambda)$  kontinuierlich auch in Sprachabschnitten erfolgen, um veränderliche Störungen innerhalb von Sprachphasen zuverlässig unterdrücken zu können. Während bei der Minimum-Statistik nach [3] Minima im zeitlichen Verlauf des Eingangs-LDS die Schätzung des Stör-LDS darstellen, wird beim rekursiven Verfahren nach [4] eine zeitliche Glättung des Eingangs-LDS mit einem variablen Glättungsfaktor durchgeführt. Da die Minimum Statistik eher zur Unterschätzung des Stör-LDS neigt und die Schätzung innerhalb eines festen Zeitfensters relativ konstant bleibt, eignet sie sich gut für Sprachabschnitte. Rekursive Verfahren liefern eine genaue Schätzung auch bei hoher Varianz des Stör-LDS und liefern daher gute Schätzergebnisse in reinen

Störabschnitten. Durch die Kombination beider Verfahren kann mit einer geringen Steigerung der algorithmischen Komplexität ein besseres Reduktionsergebnis bei stationären und leicht instationären Störungen erreicht werden.

### Kombiniertes Schätzverfahren

Abbildung 2 zeigt das Blockschaltbild des kombinierten Schätzverfahrens.

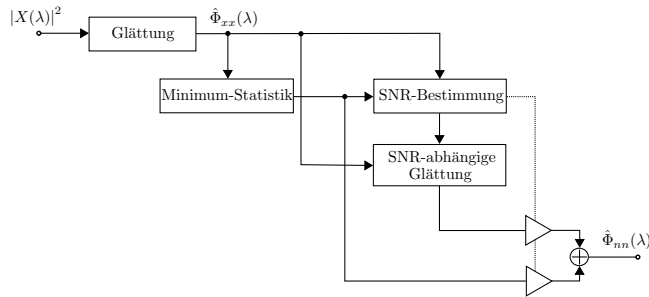


Abbildung 2: Blockschaltbild des kombinierten Algorithmus

Mit Hilfe der Minimum-Statistik wird ein maximaler SNR für jede diskrete Frequenz aus dem Eingangs-LDS geschätzt. Mit diesem SNR wird die zeitliche rekursive Glättung der Eingangsperiodogramme gesteuert. Die endgültige Schätzung setzt sich abhängig vom bestimmten maximalen SNR additiv so zusammen, dass in Sprachabschnitten das Ergebnis der Minimum-Statistik und in reinen Störabschnitten das der rekursiven Glättung überwiegt (vgl. Abbildung 3).

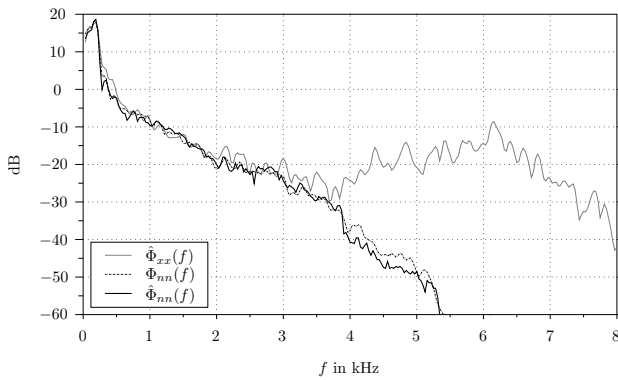


Abbildung 3: Leistungsdichtespektren

### Störgeräuschreduktion

Mit jeder Aktualisierung der Störgeräuschschätzung wird die Filterfunktion des Reduktionsfilters neu berechnet. Das parametrische Wiener-Filter der Form (1) stellt hierbei eine gute Alternative zum häufig verwendeten herkömmlichen Wiener-Filter dar.

$$H(\lambda) = \left( \frac{\hat{\xi}(\lambda)}{\alpha_{wf} + \hat{\xi}(\lambda)} \right)^{\beta_{wf}} \quad \text{mit} \quad \hat{\xi}(\lambda) = \frac{\hat{\Phi}_{ss}(\lambda)}{\hat{\Phi}_{nn}(\lambda)} \quad (1)$$

Das zur Bestimmung des a priori SNR  $\hat{\xi}(\lambda)$  erforderliche Sprach-LDS  $\hat{\Phi}_{ss}$  wird über die spektrale Subtraktion

$$\hat{\Phi}_{ss} = \hat{\Phi}_{xx}(\lambda) - \alpha_{ss} \hat{\Phi}_{nn}(\lambda) \quad (2)$$

aus dem Eingangs-LDS und dem geschätzten Stör-LDS berechnet. Da eine starke Reduktion der Störanteile das Sprachsignal zu stark verzerren kann, wird die Filterfunktion in ihrer Dämpfung begrenzt. Durch dieses so genannte Flooring bleibt ein gewisser Störanteil erhalten und das Sprachsignal wird weniger verzerrt.

### Ressourcenbedarf

Die benötigte Rechenzeit des beschriebenen kombinierten Algorithmus auf einem ADSP-BF532 von Analog Devices beträgt bei einer Abtastfrequenz von 8 kHz, einer Framelänge von 9,4 ms und einer FFT-Länge von 256 Punkten etwa 9 MIPS. Die durch den Algorithmus verursachte Verzögerung im Kanal beträgt zwei Framelängen, was bei den genannten Parametereinstellungen etwa 20 ms entspricht.

### Ergebnisse

Abbildung 4 verdeutlicht die Unterdrückung eines Fahrgeräusches beim Telefonieren im KFZ. Das vorgestellte System adaptiert sich innerhalb von 1,5 s vollständig an sprunghaft auftretende Störungen. Durch den Einfluss der Minimum-Statistik erfolgt eine sehr schnelle Adaption bei plötzlich abfallenden Störungen und die Bearbeitung reiner Sprachsignale wird kaum wahrgenommen.

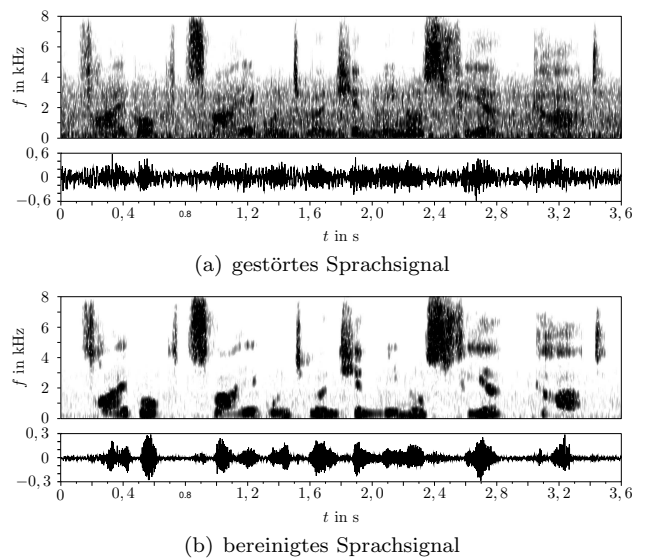


Abbildung 4: Spektrogramme zur Darstellung der Unterdrückung eines Fahrgeräusches

### Literatur

- [1] Oppenheim, A.V.; Schaffer, R.W.: Zeitdiskrete Signalverarbeitung. Pearson Studium, 2004
- [2] Loizou, P.: Speech enhancement - theory and practice. CRC Press, Taylor and Francis Group, 2007
- [3] Martin, R.: Spectral subtraction based on minimum statistics. Proc. EUSIPCO 94, pp. 1182-1185, 1994
- [4] Cohen, I.: Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Proc. Letters, vol. 9, no. 1, S.12-15, 2002