

Implementierung einer Unterwassertelefonieschnittstelle in einem SONAR-Simulator

Tim Owe Wisch¹, Till Wiegand², Florian Schulz², Gerhard Schmidt¹

¹ *Digitale Signalverarbeitung und Systemtheorie, CAU zu Kiel, E-Mail: {timw,gus}@tf.uni-kiel.de*

² *ATLAS ELEKTRONIK GmbH, Bremen, E-Mail: {till.wiegand,florian.schulz}@atlas-elektronik.com*

Kurzfassung

Sprachkommunikation unter Wasser unter Nutzung von entweder analogen, digitalen oder auch gemischten Verfahren ist noch immer ein spannendes Forschungsfeld. Durch die kostengünstige Verfügbarkeit von leistungsfähiger Hardware lassen sich die Algorithmen vollständig in Software realisieren. Im Sinne einer fortschreitenden Systemintegration ist eine zusätzliche Verwendung von Kommunikationskomponenten innerhalb von klassischen SONAR-Systemen eine logische Folge. Selbst in den Grenzen etablierter Standards können Sprachverbesserungsansätze die Verständlichkeit der Übertragung steigern und für eine verbesserte Sprachkommunikation sorgen. Aufgrund der geringen zur Verfügung stehenden Bandbreiten werden einige Sprachmerkmale, deren Energien in höheren Frequenzbereichen liegen, oftmals abgeschnitten. Während herkömmliche Techniken nur darauf abzielen, das Signal-zu-Rausch-Verhältnis zu verbessern und Spracheigenschaften außer Acht lassen, zeigt das hier vorgestellte System zusätzlich das Abbilden hochfrequenter Sprachmerkmale in niedrigere Frequenzbereiche. Auf diese Weise können auch hochfrequente Frikative über einen stark bandbegrenzten Kanal übertragen werden.

Grundlagen

Sprachkommunikation in Unterwasserumgebungen wird zumeist über Ultraschall realisiert. Durch die schwierigen Kanaleigenschaften sind die oberhalb der Wasseroberfläche eingesetzten Verfahren unter Nutzung elektromagnetischer Wellen nicht über große Entfernungen tauglich. Grundlage für die hier verwendete Implementierung ist die *STANAG* (Standardization Agreement) [1] der *NATO* (North Atlantic Treaty Organization). Diese schreibt eine Einseitenbandmodulation von Sprache vor, welche im Bereich von 300 Hz - 3 kHz übertragen wird. Diese Bandbreite ist verhältnismäßig schmal, sodass nicht alle relevanten spektralen Anteile von Sprache übertragen werden können.

In dieser Implementierung wird gezeigt, dass Signalverarbeitungsalgorithmen, die auf den Standard aufsetzen, diesen aber nicht verletzen, die Übertragung verbessern können. Das Gesamtsystem setzt sich dabei aus folgenden Teilkomponenten zusammen: Zur Unterdrückung von additiven Störungen wird eine Geräuschunterdrückung eingesetzt. Ein Kompressor am Sender reduziert die Dynamik. Zusätzlich wird am Sender und am Empfänger eine Frequenzkompression und -expansion eingesetzt, welche höherfrequente Sprachanteile in die schmale Übertragungsbandbreite einfügt und am Empfänger

breitbandigere Sprache zurückgewinnt.

Alle Algorithmen arbeiten im Frequenzbereich, sodass entsprechende Analyse- und Synthesefilterbanken am Sender und Empfänger vorgeschaltet sind [2].

Geräuschreduktion

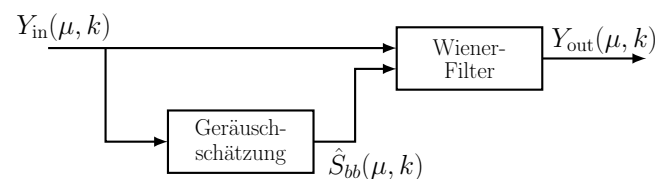


Abbildung 1: Geräuschreduktion

Die Geräuschreduktion unterdrückt störende Hintergrundgeräusche und besteht aus zwei Teilen: Der erste Teil (Geräuschschätzung [3]) ist die Schätzung des Kurzzeitleistungsdichtespektrums des Hintergrundgeräusches $\hat{S}_{bb}(\mu, k)$. Der zweite Teil die Entfernung dessen vom Eingangssignal mittels eines Wiener-Filters. Diese Aufteilung ist in **Abbildung 1** schematisch dargestellt. Für die Geräuschschätzung wird zunächst das Signal auf Sprachaktivität überprüft, damit in die Schätzung nur Hintergrundgeräusche eingehen. Zusätzlich wird ein Trend der Störgeräusche geschätzt um bei schnellen Änderungen der Umgebungsparameter die Schätzung entsprechend anpassen zu können. Mit dem geschätzten Geräuschspektrum wird durch ein Wiener-Filter das Eingangssignal vom Rauschen befreit. Für die Übertragungsfunktion des Wiener-Filters gilt [4]:

$$H_{\text{opt}}(\mu, k) = \max \left\{ H_{\text{min}}, 1 - \frac{\hat{S}_{bb}(\mu, k)}{\hat{S}_{yy}(\mu, k)} \right\}, \quad (1)$$

wobei $\hat{S}_{bb}(\mu, k)$ das geschätzte Leistungsdichtespektrum der additiven Störung bezeichnet und $\hat{S}_{yy}(\mu, k)$ das Kurzzeitleistungsdichtespektrum des gestörten Eingangssignals.

Kompressor

Der Kompressor am Sender minimiert den Pegelunterschied zwischen lauten und leisen Passagen eines Eingangssignals [5]. Die Attack- und Releaseparameter definieren Zeitkonstanten, die angeben, wie schnell der Kompressor ab dem Schwellwert anfängt zu regeln bzw. die Regelung abschaltet (siehe **Abbildung 3**).

Modulation

Die spektrale Verschiebung des Signals wird im Frequenzbereich ausgeführt, indem die im betreffenden Frequenz-

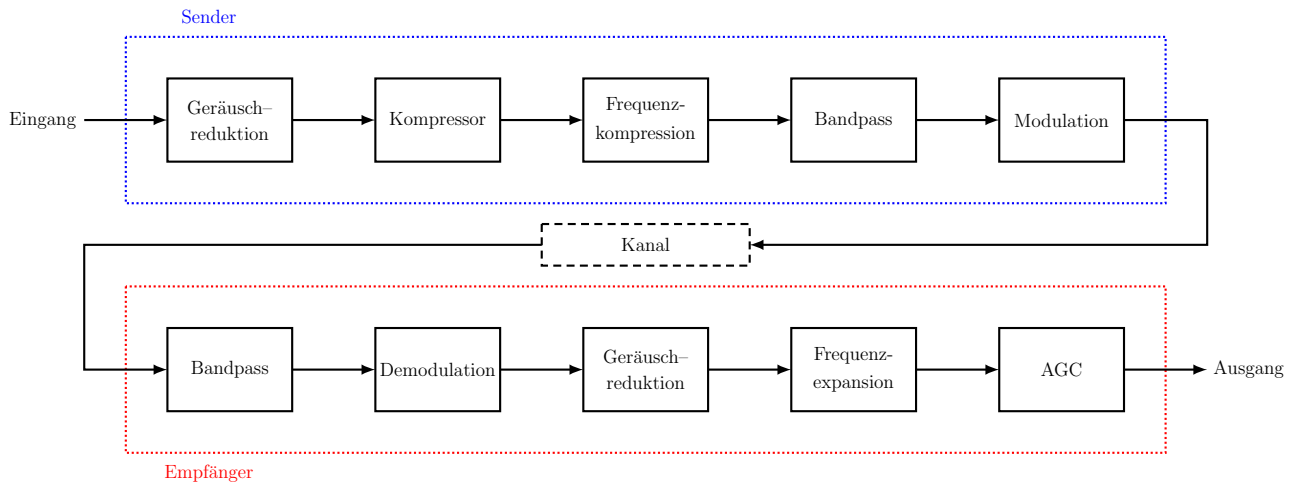


Abbildung 2: Modulübersicht der Sende- und Empfangsstrecke

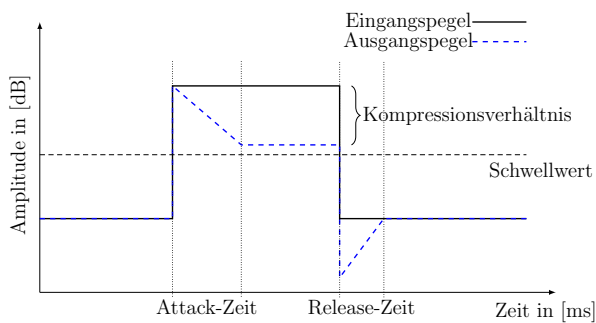


Abbildung 3: Regelverhalten des Kompressors

bereich liegenden FFT-Bins verschoben werden:

$$\text{DFT} \left\{ x(n) \cdot e^{j\mu_0 \frac{2\pi}{M} n} \right\} = X(\mu - \mu_0). \quad (2)$$

Durch die gewählte Struktur mit Analyse- und Synthesefilterbanken und Overlap-Add Verfahren ist noch ein zusätzlicher Phasenterm nötig, damit die Verschiebung verzerrungsfrei durchgeführt werden kann. Insgesamt ergibt sich für die Frequenzverschiebung:

$$X_{\text{mod}}(\mu) = X(\mu - \mu_0) \cdot e^{j f_{\text{mod}} \cdot 2\pi \cdot \frac{R}{f_s}}. \quad (3)$$

In Gleichung (3) bezeichnet μ den diskreten Frequenzindex, R den Rahmenversatz, f_s die Abtastfrequenz und f_{mod} die gewünschte Modulationsfrequenz.

Automatische Verstärkungsregelung

Die automatische Verstärkungsregelung (engl. *Automatic Gain Control [AGC]*) am Empfänger sorgt für eine angemessene Aussteuerung des empfangenen Signals. Wichtig zu beachten ist dabei allerdings, dass das Rauschen nicht angehoben werden sollte. Zu diesem Zweck besitzt die vorgestellte AGC einen eigenen Geräuschschätzer, sodass die Verstärkung bzw. Dämpfung nur nachgeregelt wird, wenn ein gewisses SNR vorhanden ist und Sprachaktivität detektiert wird.

Frequenzkompression

Die durch den Standard geforderte stark limitierte Bandbreite (300 Hz bis 3 kHz) erschwert eine gute Verständlichkeit bei der Verbindung. Kommen zusätzlich noch Kanaleinflüsse hinzu, sinkt die Verständlichkeit weiter. Die Frequenzkompression [6] und -expansion hat das Ziel Spektralanteile in der Sprache, die eigentlich außerhalb der Bandbegrenzung liegen, in das Sendesignal zu integrieren (ein breiteren Frequenzbereich in einen niedrigeren zu komprimieren) und im Empfangssignal wieder eine Rekonstruktion (Expansion) durchzuführen. Dieses Verfahren ist jedoch nicht verlustlos und bei Vokalen ist die Verständlichkeit trotz der geringen Bandbreite kaum eingeschränkt. Aus diesem Grund wird dieses Verfahren nur auf Frikative angewandt (z.B. s-Laute). Frikative haben hohe Energieanteile außerhalb der geforderten Bandbreite, sodass diese ohne Kompression kaum übertragen werden können.

Bei der Kompression werden die niedrigeren Frequenzen nicht beeinflusst, da das Eingangssignal in ein höheres und tieferes Frequenzband aufgeteilt und für die Kompression nur das obere Frequenzband genutzt wird. Die Grenze zwischen den beiden Bändern liegt bei 2000 Hz. Die Detektion eines Frikativs erfolgt über einen Vergleich der Energien zwischen den beiden Frequenzbändern und einer Abschätzung des Rauschgehalts, da Frikative keine Vielfachen der Sprachgrundfrequenz aufweisen. Je nach verwendeter Kennlinie werden auf den Bereich von 2 kHz bis 3 kHz höhere Frequenzbereiche aufgeprägt. Beispielsweise bei einer Steigung der Kennlinie von 2 Frequenzen bis 4 kHz. Um eine Kompression auf das Eingangssignal anzuwenden, wird dieses zunächst geglättet. Die Glättung erfolgt frequenzabhängig mit folgenden Zeitkonstanten:

$$\Delta_{\text{smooth}}(\mu) = 2595 \cdot \log_{10} \left(\frac{L_{\mu, f}}{700 \text{ Hz}} + 1 \right) - 2595 \cdot \log_{10} \left(\frac{L_{\mu-1, f}}{700 \text{ Hz}} + 1 \right). \quad (4)$$

Deutlich zu erkennen ist, dass es sich hier um eine auf der Mel-Skala beruhende Glättung handelt. Geglättet wird durch:

$$\overline{Y_{\text{smooth}}(\mu, k)} = \Delta_{\text{smooth}}(\mu) \cdot \overline{Y_{\text{smooth}}(\mu, k-1)} + (1 - \Delta_{\text{smooth}}(\mu)) \cdot Y_{\text{in}}(\mu, k). \quad (5)$$

Mithilfe des geglätteten Eingangsspektrums werden die einzelnen FFT-Bins oberhalb einer Grenzfrequenz μ_{min} nun komprimiert nach:

$$Y_{\text{out}}(\mu, k) = \frac{Y_{\text{in}}(\mu, k)}{|\overline{Y_{\text{smooth}}(\mu, k)}|} \cdot \sum_{i=\mu_{\text{low}}}^{\mu_{\text{high}}} W_i \cdot Y_{\text{in}}(\mu_i, k). \quad (6)$$

Da bei der Kompression mehrere Eingangsbins auf einen Ausgangsbin gemappt werden, müssen die Eingänge entsprechend ihrem Anteil am Ausgang gewichtet werden (W_i). Dazu wird linear interpoliert:

$$y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} \cdot (x - x_1). \quad (7)$$

Die Bestimmung der Gewichte folgt aus der resultierenden Nachkommastelle der Interpolation. Für die Gewichte gilt:

$$\sum_i W_i = 1. \quad (8)$$

Expansion

Die Expansion ist die inverse Operation zur Kompression. Zu beachten ist allerdings, dass die Kompression nicht verlustlos arbeiten kann, da eine Rekonstruktion der zusammengeschobenen Bins nicht mehr ohne Zusatzinformation möglich ist. Da jedoch keine zusätzlichen Informationen vorliegen, wird versucht das ursprüngliche Spektrum so gut wie möglich zu rekonstruieren. Wie bei der Kompression wird das Eingangssignal zunächst in einen höheren und einen niedrigeren Frequenzanteil aufgeteilt:

$$Y_{\text{out}}(\mu, k) = \frac{Y_{\text{in}}(\mu, k)}{|\overline{Y_{\text{smooth}}(\mu, k)}|} \cdot \sum_{i=\mu_{\text{low}}}^{\mu_{\text{high}}} W_i \cdot \overline{\overline{Y_{\text{smooth}}(i, k)}}. \quad (9)$$

Die Expansion wird, analog zur Kompression, bei erkanntem Frikativ durchgeführt. Der einzige Unterschied besteht in dem doppelt geglätteten Eingangsspektrum ($\overline{\overline{Y_{\text{smooth}}(\mu, k)}}$), welches gleichzeitig als Normierung und Anregung genutzt wird.

Softwarestruktur

Die Einzelmodule und deren Reihenfolge ist in **Abbildung 2** dargestellt. Zunächst wird auf das Eingangssignal eine Geräuschreduktion angewendet, damit möglichst nur Sprache ohne Hintergrundgeräusche

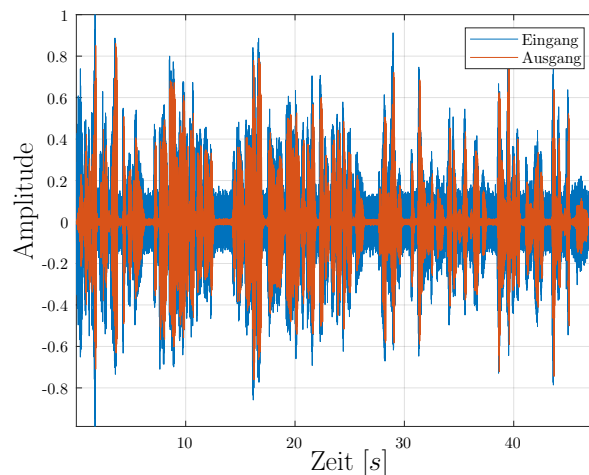


Abbildung 4: Eingangs- und Ausgangssignal der Geräuschunterdrückung

übertragen wird. Darauf folgt ein Kompressor, welcher den Dynamikbereich des Eingangssignals minimiert. Die Frequenzkompression komprimiert Frequenzanteile des breiten Eingangsfrequenzbereichs in den schmalen Übertragungsbereich des Standards. Darauf wird die Bandpassfilterung und Modulation angewendet. Am Empfänger folgt nach Filterung und Demodulation wieder eine Geräuschreduktion, die Expansion der komprimierten Frequenzen und anschließend eine automatische Aussteuerungsregelung.

Ergebnisse

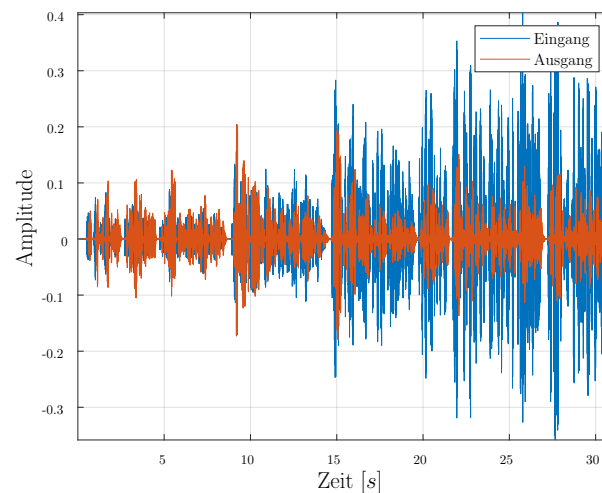


Abbildung 5: Verstärkungsregelung bei einem im Pegel ansteigenden Signal

Die Geräuschreduktion trennt bei korrekter Schätzung der Störung ein Nutzsinal von einem Störsignal. In dieser Simulation wurden einerseits der Einfluss von weißem Rauschen auf das Signal untersucht, andererseits Unterwassergeräusche, die in dem Frequenzbereich der

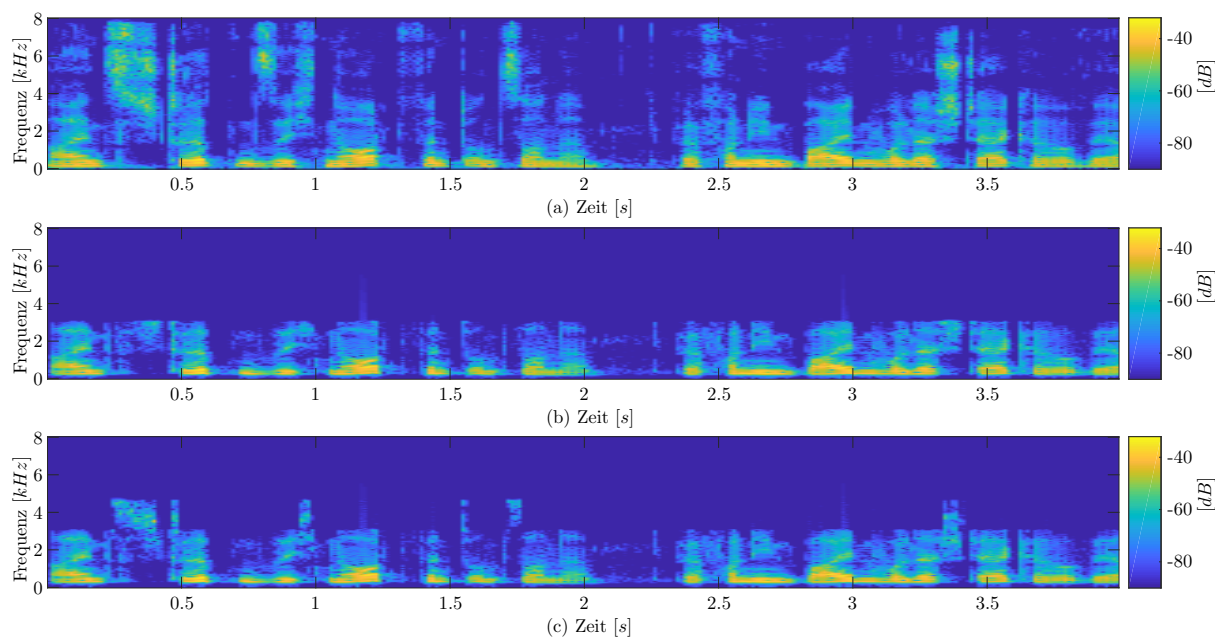


Abbildung 6: (a) Originalsignal, (b) bandbegrenzt Signal und (c) Frequenzexpansion

Übertragung liegen auf das Signal addiert.

Die AGC sorgt für eine passende Aussteuerung des Signals am Empfänger. Da der Eingangspegel über eine größere Entfernung absinken kann, ist dieser algorithmische Schritt notwendig um ein hörbares Ausgangssignal zu generieren. Getestet wird dies mithilfe eines Signals, welches in mehreren Schritten im Pegel steigt. Im Idealfall steuert die AGC dies nach, indem das Signal in schwachen Passagen verstärkt wird und in lauten Passagen abgesenkt wird. Dieser Steuerungsprozess ist in **Abbildung 5** abgebildet.

In **Abbildung 6** sind die Resultate der Frequenzkompression bzw. der Expansion zu sehen. Das in der Abbildung oben dargestellte Spektrum zeigt, dass oberhalb von 3 kHz an mehreren Stellen Leistung vorhanden ist. Diese Signalanteile würden normalerweise weggefiltert und nicht übertragen werden. Die Frequenzkompression komprimiert diese Signalanteile, sodass das Sendesignal, welches im mittigen Spektrum dargestellt ist, dem Standard entsprechend nur Anteile bis 3 kHz enthält. Nach der Modulation, Übertragung, Empfang und Demodulation steht am Empfänger wieder ein Signal mit der im Standard vorgeschriebenen Bandbreite zur Verfügung. Die eine Möglichkeit besteht nun darin, das empfangene Signal ohne Expansion wieder auszugeben, jedoch liegen die komprimierten Frikative in einem zu tiefen Frequenzbereich. Um diese wieder in die ursprünglichen Bereiche zu verschieben wird auf das Signal die Expansion angewendet. Die rekonstruierten Anteile oberhalb von 3 kHz bis 4,5 kHz sind im unteren Spektrum deutlich zu erkennen (z.B. von 300 ms bis 500 ms). Da keine Randinformationen übertragen werden, kann das Verfahren nicht verlustlos funktionieren. Die aus dem niedrigeren Bändern rekonstruierten Frequenzen weisen lediglich ei-

ne Struktur auf, die ähnlich zu dem Originalsignal ist. Aus diesem Grund wird das Verfahren auch nur bei Frikativen angewandt, da bei stimmhaften Lauten die unregelmäßige Struktur der Pitchlinien auffallen würde. Die Frequenzkompression und -expansion kann so dafür sorgen, dass auch scharfe s-Laute über den Unterwasserkanal gemäß dem Standard übertragen werden können und bildet so zusammen mit vorher genannten Erweiterungen eine deutliche Verbesserung für die Verständlichkeit bei der Unterwassertelefonie.

Literatur

- [1] NSO: *NATO Standardization Office*, <https://www.nsa.nato.int/nsa/>, Abgerufen 18.11.2019
- [2] A. V. Oppenheim, R. W. Schaffer, J. R. Buck: *Discrete-Time Signal Processing*, Prentice Hall, 1999
- [3] V. Kandade Rajan: *Speech Enhancement in Hands-free Systems for Automobile Environments*, Dissertation, Shaker Verlag, 2017
- [4] Y. Huang, J. Benesty, M. M. Sondhi (Eds.): *Springer Handbook of Speech Processing*, Springer Verlag, 2008
- [5] U. Zölzer: *DAFX-Digital Audio Effects*, Wiley, New York, 2002
- [6] J. Withopf, P. Hannon, M. Krini, G. Schmidt: *Phoneme-Dependent Speech Enhancement*, 9th ITG-Symposium Speech Communications (ITG '10), VDE, Bochum, Deutschland, Sept. 2010