

Nichtlineare Signalverarbeitung für Systeme zur Unterwassertelefonie

Tim Owe Wisch¹, Gerhard Schmidt¹

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

Kurzfassung

Sprachkommunikation unter Wasser unter Nutzung von sowohl analogen, digitalen und auch gemischten Verfahren ist noch immer ein spannendes Forschungsfeld. Durch die breite Verfügbarkeit von leistungsfähiger Hardware entsteht die Möglichkeit, sich die Vorteile etablierter Verfahren in Kombination mit neuen, komplexen Algorithmen zu Nutzen zu machen. Durch die physikalisch bedingten Begrenzungen unter Wasser und im Hinblick auf die spektrale Effizienz kann bspw. die bekannte Technik der Einseitenbandmodulation ohne Träger verwendet werden. Auf der Sprachsignalseite können in einem solchen System verschiedene Signalanpassungen vorgenommen werden, um die Ansteuerung der analogen Verstärker zu verbessern und die Sendeleistung der Schallwandler optimal auszunutzen. Zudem können verschiedene Limiter, Kompressoren und weitere Verfahren angewendet werden, um die Robustheit des Systems gegenüber Störungen im Wasser und Hardwarebeschränkungen zu verbessern. Zur Evaluation der Auswirkungen der Verfahren auf das Sprachsignal werden subjektive und objektive Tests (STOI-Algorithmus) verwendet.

Grundlagen

Für Sprachkommunikation unter Wasser wurden in den vergangenen Jahren viele Systeme zur digitalen Übertragung von Sprache vorgestellt. Ursprünglich wurden analoge Verfahren verwendet, welche jedoch unter den Gegebenheiten des Unterwasserkanals oftmals Unzulänglichkeiten zeigten. Andererseits sind auch digitale Systeme nicht von Fehlfunktionen befreit: Während unter ungünstigen Kanalbedingungen die analoge Kommunikation zwar gestört, in Teilen jedoch noch verständlich sein kann, führt eine zu hohe Bit-Error-Rate (BER) bei einer digitalen Übertragung zwangsläufig zu einer abbrechenden Verbindung. Während man bei digitalen Übertragungen auf niederbitratige Codecs und aufwendige Fehlerkorrektur zurückgreifen muss, kann diese Aufgabe bei analoger Übertragung im weitesten Sinne vom menschlichen Gehör übernommen werden. Das Gehirn ist durchaus in der Lage mit vielen verschiedenen Rauschszenarien umzugehen, sodass bei einer analogen Übertragung eher eine Art Hilfestellung für das Gehör mit psychoakustischen Effekten gegeben werden kann. Ein Kommunikationssystem sollte in erster Linie in der Lage sein die übergebenen Informationen mit bestmöglicher Qualität und Verständlichkeit zu übertragen. Die Einschätzung der Sprachverständlichkeit lässt sich entweder mit Hilfe von Testpersonen bestimmen, oder durch Algorithmen automatisiert schätzen.

Ein solcher Algorithmus ist beispielsweise der STOI (Short-Time Objective Intelligibility)-Algorithmus [1]. Dabei wird mit Hilfe von gewichteten Zeit-Frequenz-Analysen die Ähnlichkeit von gestörter und klarer Sprache ermittelt.

Unterwassertelefonie

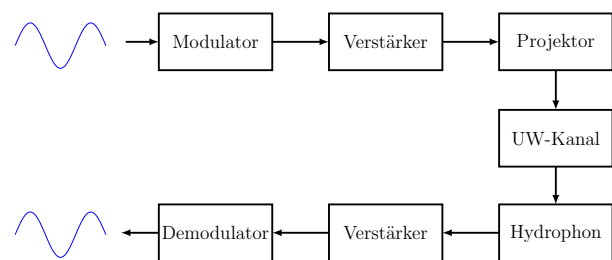


Abbildung 1: UW-Telefon Systemaufbau.

Im Bereich analoger Unterwassertelefonie wird in kommerziellen Produkten häufig die Einseitenbandmodulation genutzt [2], da sie sowohl einfach umzusetzen ist, als auch eine gute spektrale Effizienz bietet. Die Bandbreite des gesendeten Signals entspricht dabei der Bandbreite des (bandbegrenzten) Sprachsignals. Die Demodulation lässt sich, obwohl komplexer als Zweiseitenbandmodulation, ebenfalls mit Hilfe von günstiger Hardware realisieren, sodass insgesamt ein verhältnismäßig einfaches wie günstiges System entsteht. Digitale Systeme leiden in vielen Fällen unter der niedrigen Datenrate, sodass nur Codecs mit niedrigen Bitraten und damit verhältnismäßig schlechter Qualität verwendet werden können. Ein großer Vorteil digitaler Systeme ist jedoch die einfachere Fehlererkennung und Fehlerkorrektur. Obgleich inzwischen digitale Systeme zur Verfügung stehen, wird in diesem Beitrag zunächst auf die Möglichkeiten zur Optimierung analoger Systeme eingegangen. Bei der direkten Verwendung von Sprache als Eingangssignal kann sich die hohe Dynamik eines Sprachsignals negativ auf die Übertragungseigenschaften auswirken. Die Dynamik eines Sprachsignals kann bis zu 50 dB [3] betragen, was dazu führt, dass sowohl kleine, als auch sehr große Amplituden im Signal vorkommen. Für die Sprachverständlichkeit sind alle Werte von Bedeutung. Bei der Übertragung über einen stark gestörten Unterwasserkanal kommen neben Mehrwegeausbreitung auch je nach Frequenzbereich mehr oder weniger starkes additives Rauschen hinzu (Schiffsverkehr, Wetterbedingungen, etc.). Dieses Rauschen maskiert kleinere Signalamplituden und kann am Empfänger nur eingeschränkt entfernt werden. Eine Möglichkeit besteht darin, den

Verstärkungsfaktor der Sendeverstärker zu erhöhen, sodass auch kleine Signale sich aus dem Rauschen abheben. Nachteilig ist dabei allerdings, dass Verstärker bei zu großen Signalamplituden nichtlinear werden, bzw. es zu Clipping mit den damit einhergehenden Verzerrungen kommt, und nur bei einzelnen Signalspitzen voll ausgesteuert werden. Aus diesem Grund ist es sinnvoll, vor der Übertragung eine Dynamikreduktion oder eine ähnliche Vorverarbeitung des Sprachsignals vorzunehmen, um eine einheitlichere Aussteuerung der Verstärker zu gewährleisten.

Signalmodifikationen

Zur verbesserten Aussteuerung der Verstärker und verbesserten Sprachverständlichkeit am Empfänger stehen verschiedene Signalmodifikationen zur Verfügung, die im Folgenden genauer betrachtet werden.

Limiter

Eine sehr einfache Möglichkeit zur verbesserten Aussteuerung der Verstärker für die Unterwassertelefonie besteht in der Verstärkung des Signals und anschließendes Begrenzen in Software, sodass die Verstärker nicht in Sättigung gehen. Die zum Begrenzen eingesetzten Limiter ([5], [6]) können einerseits eine harte Charakteristik aufweisen (Gl. 1), oder weicher in die Sättigung gehen, beispielsweise durch eine quadratische Kennlinie (Gl. 2). Beide Limiterarten bringen unweigerlich Verzerrungen mit sich, deren Auswirkungen im folgenden Abschnitt weiter beleuchtet werden.

$$h(x) = \begin{cases} -1 & , x \leq -1 \\ x & , -1 < x < 1 \\ 1 & , x \geq 1 \end{cases} \quad (1)$$

$$l(x) = \begin{cases} -1 & , x \leq -1 \\ \frac{3x}{2} \left(1 - \frac{x^2}{3}\right) & , -1 < x < 1 \\ 1 & , x \geq 1 \end{cases} \quad (2)$$

Kompressor

Eine aus der üblichen Telefonie bekannte Technik ist das μ -Law Quantisierungsverfahren [7]. Bei diesem Verfahren werden die Eingangssignale über eine logarithmische Kompressor-Kennlinie verzerrt (Gl. 3), sodass sich der Dynamikbereich verkleinert und dadurch AD-Wandler mit weniger Bits verwendet werden können.

$$c(x) = \text{sgn}(x) \left(\frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \right) \quad (3)$$

Am Empfänger kann dies durch die zum Kompressor inverse Expander-Kennlinie (Gl. 4) rückgängig gemacht werden, sodass wieder die volle Dynamik zur Verfügung steht.

$$c^{-1}(x) = \text{sgn}(x) \left(\frac{(\mu + 1)^{|x|} - 1}{\mu} \right) \quad (4)$$

In Abb. 2 ist ein Sprachsignal vor und in Abb. 4 nach dem Kompressor abgebildet. Das Verstärken von kleinen Pegeln und das verbesserte Aussteuern durch die reduzierte Dynamik ist klar erkennbar.

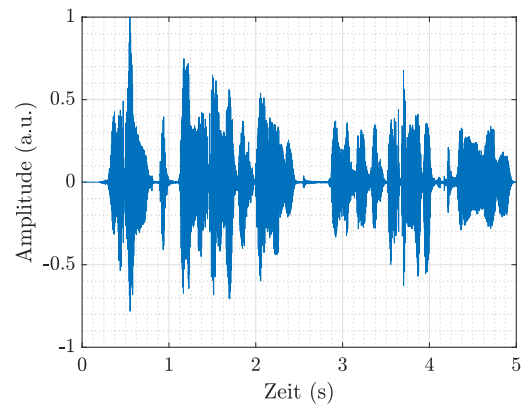


Abbildung 2: Sprachsignal ohne Dynamikreduktion.

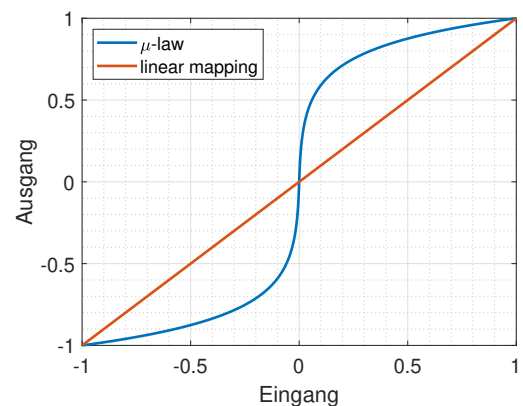


Abbildung 3: Nichtlineares μ -law Mapping.

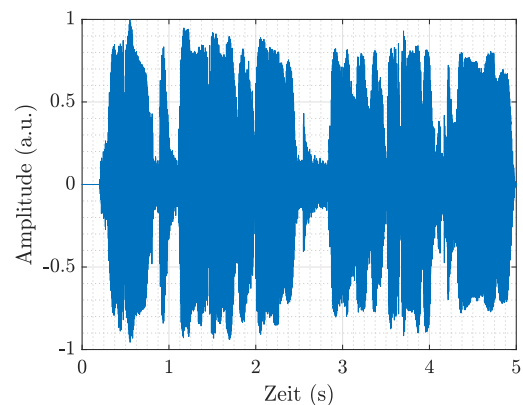


Abbildung 4: Sprachsignal mit Dynamikreduktion durch μ -law Mapping.

Multiband Enhancer

Ein weiteres Instrument zur Verbesserung der Sprachverständlichkeit am Empfänger besteht in der Nutzung von verschiedenen Nichtlinearitäten (Abb. 5), um in verrauschten Signalen die Harmonischen eines Sprachsignals herauszuheben. Dazu wird das Eingangssignal am Empfänger in fünf verschiedene Bänder verzweigt. Band 0 ist ein direkter Durchgriff zum Ausgang, die Bänder 1-4 werden bandbegrenzt und durch je nach Band mit höherwertigen Exponenten versehen. Mit den Gewichten w_0 bis w_4 kann der Einfluss des Durchgriffs und der ein-

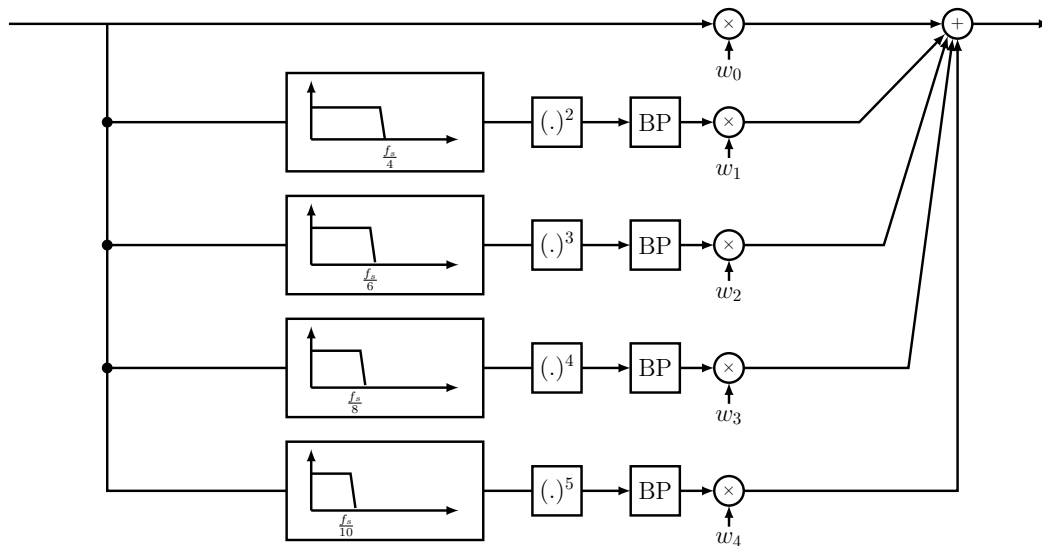


Abbildung 5: Multiband Enhancer nach [8].

Band	Gewicht w_i	Grenzfrequenz BP
0: $\frac{f_s}{2}$	0.5	-
1: $\frac{f_s}{4}$	0.1	50-800 Hz
2: $\frac{f_s}{6}$	0.3	50-600 Hz
3: $\frac{f_s}{8}$	0.05	50-400 Hz
4: $\frac{f_s}{10}$	0.1	50-200 Hz

Tabelle 1: Gewichte und Grenzfrequenzen des Enhancers.

zelenen Bänder geregelt werden. Die Bandpässe dienen zur Unterdrückung von Gleichanteilen. Am Ausgang werden alle Bänder zusammenaddiert. In (Tab. 1) sind die simulativ ermittelten Gewichte und Grenzfrequenzen aufgeführt.

Simulation und Ergebnisse

Um den Einfluss der verschiedenen Signalmodifikationen zu evaluieren, werden die Signale in einem ersten Schritt mit verschiedenen Signal-zu-Rausch-Verhältnissen versehen und die Sprachverständlichkeit mit dem STOI-Algorithmus ermittelt. Der STOI-Wert gibt dabei die prädierte Worterkennungsrates an. Es gilt zu beachten, dass Sprachpausen bei der Berechnung der Signalenergie nicht berücksichtigt werden, sodass eine realistische Einordnung möglich wird. Die SNR-Werte für die Signalmodifikationen werden immer bezogen auf das unprozessierte Signal angegeben. Für die Sprachdaten wird auf verschiedene Sprecher aus der NTT-Datenbank [4] zurückgegriffen. In Abb. 6 wird deutlich, dass bei einem unprozessierten Signal die Sprachverständlichkeit mit niedrigerem SNR stark abnimmt.

Der Einsatz von Limitern am Sender bzw. Kompressor/Expander-Paaren am Sender und Empfänger führt in diesem einfachen Simulationsfall zu einer deutlichen Verbesserung und zu einem flacheren Abfall der Verständlichkeitskurve. Für die Limiter wurde das Signal um den Faktor 3 verstärkt und

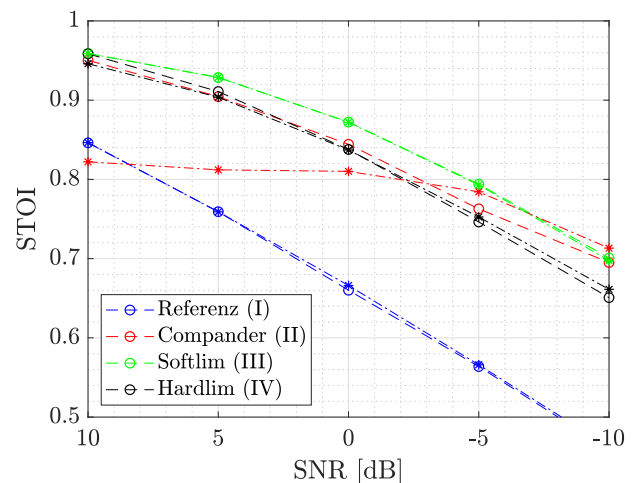


Abbildung 6: STOI-Ergebnis über SNR ohne Mehrwegeausbreitung.

danach der jeweiligen Begrenzercharakteristik unterworfen. Während die Verständlichkeit eines unprozessierten Signals bei -5 dB SNR bereits deutlich beeinträchtigt und nur noch knapp unter 60 % beträgt, kann das prozessierte Signal noch 75 % erreichen. In gleicher Farbe aber mit *-Symbol sind die Ergebnisse mit dem zusätzlichen Einsatz eines Multiband-Enhancers gekennzeichnet.

Zusätzlich erfolgt eine Simulation mit dem unter Wasser häufig vorkommenden Phänomenen der Mehrwegeausbreitung, wobei zwischen Sender und Empfänger neben dem direkten Pfad zwei weitere Pfade mit 20 und 60 ms Verzögerung und unterschiedlichen Phasendrehungen angenommen werden. Diese zusätzlichen Pfade führen zu einer weiteren Verschlechterung der Verständlichkeit, da neben Echo- auch Interferenzeffekte auftreten. Beiden Simulationen gemein ist die Erkenntnis, dass die Sprachverständlichkeit durch nichtlineare Kennlinien im Signalweg gesteigert werden kann. Bereits bei 5 dB SNR liegt das unprozessierte Signal in der Verständlichkeit

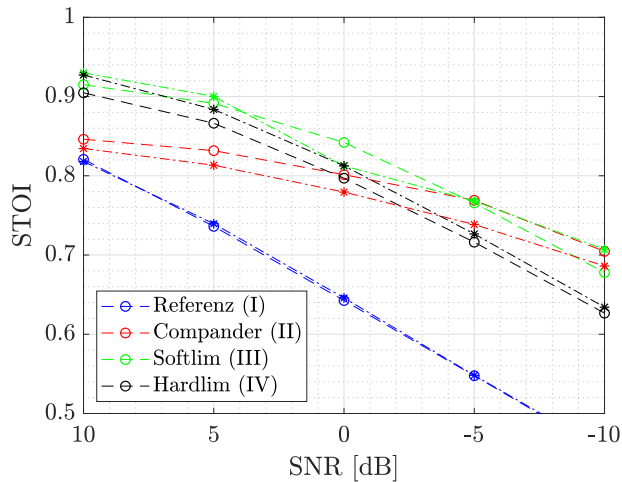


Abbildung 7: STOI-Ergebnis über SNR mit Mehrwegeausbreitung.

deutlich unter den Limitern und dem Compander. Besonders der Softlimiter kann sich gegenüber den anderen Verfahren behaupten und wird erst bei -10 dB SNR vom Compander eingeholt, bzw. im Fall der Mehrwegeausbreitung sogar überholt. Bemerkenswert ist die verhältnismäßig schlechte Performance des Companders bei gutem SNR im Mehrwegefall, bei dem kaum Verbesserungen gegenüber dem unprozessierten Signal auftreten. Zusätzlich zu den betrachteten Nichtlinearitäten wird noch der Multiband-Enhancer am Empfänger betrachtet. Die Simulationsergebnisse sind in diesem Fall indifferent, da beispielsweise der Hardlimiter zu profitieren scheint, der Softlimiter jedoch kaum Vorteile erhält. Beim Compander wird das Ergebnis sogar bei guten SNR-Werten verschlechtert. Im einfachen Simulationsfall wird dieser Verlust zwar durch eine Verbesserung bei niedrigen SNR kompensiert, bei angenommener Mehrwegeausbreitung kann diese Verbesserung jedoch nicht nachgewiesen werden. Insgesamt zeigen alle Nichtlinearitäten im Signalweg eine Verbesserung gegenüber dem unprozessierten Signal. Da in einer realitätsnahen Anwendung das SNR nicht ohne weiteres im Vorhinein bekannt sein kann und sich durchaus während des Betriebes ändern kann, ist insgesamt der Softlimiter die beste Wahl. In der Simulation wird die Performance lediglich bei sehr schlechten SNR-Werten vom Compander übertroffen. Der Einsatz des Multiband-Enhancers erzielt nicht den erhofften Erfolg und führt nur in einigen Fällen zu leichten Verbesserungen, kann andererseits aber auch in ungünstigen Kombinationen zu einer Verschlechterung der Performance führen.

Zusammenfassung und Ausblick

In diesem Beitrag wurde gezeigt, dass verschiedene Signalmodifikationen durchaus in der Lage sind die Sprachverständlichkeit zu verbessern. Die Implementierung von Kompressor-Expander-Paaren in Sender und Empfänger erfordert keine rechenaufwendigen Algorithmen und lässt sich sogar kostengünstig in Hardware mit bereits verfügbaren Chips realisieren. Wird das Sprachsi-

gnal jedoch digitalisiert, ist eine alternative Implementierung zB. eine Verstärkung mit anschließendem Limiter in Software möglich. Im lehrstuhleigenen Echtzeitrahmenwerk KiRAT [9] ist aus diesem Grund aktuell eine Limiterimplementierung für das Unterwassertelefon vorhanden und wird im Laufe des Jahres in der Ostsee und Kieler Förde ausgiebig getestet werden. Da der vorgestellte Multiband-Enhancer laut Simulation keine größeren Verbesserungen brachte, wird auf eine Implementierung verzichtet. Insgesamt sind die Möglichkeiten analoger Telefonie unter Wasser jedoch begrenzt und die mangelnde Fehlerkorrektur sorgt dafür, dass eine gute Anpassung an Umweltbedingungen nur sehr eingeschränkt möglich ist. Digitale Ansätze versprechen Lösungen für viele Probleme, kranken jedoch an dem unter Wasser stetig vorherrschenden Problem geringer Bandbreiten und Datenraten. Eine Kombination digitaler und analoger Ansätze könnte beiden Ansätzen für sich überlegen sein. Beispielsweise könnte eine Reduktion der Dynamik nicht wie hier vorgestellt über Compander oder Limiter erfolgen, sondern stattdessen das Signal in einzelne Blöcke aufgeteilt und adaptiv angepasst werden. Die nötigen Daten zur Rückgewinnung der Sprachdaten müssten dann auf einem digitalen Kanal übertragen und am Empfänger decodiert werden. Ebenfalls denkbar ist eine Übertragung der spektralen Einhüllenden des Sprachsignals, sodass Einflüsse durch den Kanal minimiert werden können. Zusätzlich bestünde die Möglichkeit in Sprachpausen bekannte Sequenzen zu senden, um mit den daraus gewonnenen Daten Equalizer zu parametrieren und Kanaleinflüsse weiter zu minimieren.

Literatur

- [1] C.H. Taal et al.: An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech, IEEE Transactions on Audio, Speech and Language Processing, 2011
- [2] Ocean Technology MkII Buddy Phone, <https://www.leisurepro.com/p-otsmkiibud/ocean-technology-mkii-buddy-phone>
- [3] B. Schneider, W. Bigenzahn: Stimm diagnostik, Springer, Wien, 2007
- [4] NTT-AT: Multilingual Speech Database. NTT Advanced Technology Corporation, Tokyo, 2002
- [5] G. Ballou: Handbook for Sound Engineers, Focal Press, Waltham, 1987
- [6] T. Araya, A. Suyama: Sound effector capable of imparting plural sound effects like distortion and other effects, US Patent 5,570,424, 29 Oct. 1996
- [7] ITU-T Recommendation G.711
- [8] U.Zölzer: DAFX-Digital Audio Effects, Wiley, New York, 2002
- [9] Kiel Realtime Audio Toolkit, <https://dss.tf.uni-kiel.de/index.php/research/realtime-framework/kielat-overview>