Geräuschreduktion und Entzerrung für gemischt analog-digitale Sprachübertragungen im Unterwasserkanal

Tim Owe Wisch[1], Bastian Kaulen[1], Frederik Kühne[1], Gerhard Schmidt[1]
[1] *CAU Kiel, 24143 Kiel, Deutschland, Email: {timw, bk, frk, gus}@tf.uni-kiel.de*

Einleitung

In vielen Anwendungsbereichen ist unter Wasser eine Sprachkommunikationsverbindung zwischen zwei Partnern erforderlich. Klassischerweise werden dazu entweder die direkte Einseitenbandmodulation eines Sprachsignals oder eine digitale Kommunikationsform unter Nutzung eines Sprachcoders angewandt. Die besonderen Bedingungen des Unterwasserkanals sorgen jedoch dafür, dass einerseits das direkt modulierte Sprachsignal starken Verzerrungen im Kanal unterworfen wird und andererseits digitale Kommunikationsverbindungen oftmals nur geringe Datenraten liefern können und demzufolge nur Sprachcoder mit niedriger Qualität eingesetzt werden können. Eine mögliche Alternative besteht in dem Ansatz ein System einzusetzen, welches in einem gemischten Betrieb sowohl digitale als auch analoge Übertragungen kombiniert. Im digitalen Teil werden Informationen über die Einhüllende und ein Normalisierungsfaktor unter Verwendung von Codebüchern (Vektorquantisierung), im analogen Teil ein spektral ausgewogenes und normalisiertes Restsignal übertragen. Ein solches System wird in diesem Beitrag in Bezug auf die gegenüber den klassischen Varianten erweiterten Möglichkeiten zur Geräuschreduktion und Entzerrung untersucht und evaluiert.

Grundlagen

Fast alle digitalen Sprachcodierer schätzen bestimmte Sprachmerkmale, um das Sprachsignal zu modellieren und eine gute Darstellung des Eingangssignals zu finden. Dies wird mit einer Datenkompression kombiniert, um einen kompakten Bitstrom zu erzeugen. Eine weit verbreitete Technik ist *linear predictive coding* (LPC), welche die spektrale Hüllkurve eines Sprachrahmens schätzt. Die LPC-Filterung führt zu einer Aufteilung des Signals in Anregung, bzw. Restsignal, und Einhüllende. Beide werden im digitalen Fall für die Übertragung quantisiert. Für die Sprachkommunikation unter Wasser wird oftmals ein analoger auf Einseitenbandmodulation basierender Ansatz verwendet, bei dem keine digitalen Elemente genutzt werden (siehe **Abbildung 1**).

Traditioneller Ansatz

Der traditionelle, vom *Standardization Agreement* der NATO (STANAG) definierte Ansatz erfordert einen Bandpassfilter für die Eingangssprache $s(n)$, der den Eingangsfrequenzbereich auf 300 Hz - 3 kHz begrenzt. Anschließend wird das Sprachsignal durch Einseitenbandmodulation auf eine Frequenz von 8087,5 Hz gemischt [1]. Kommerzielle Produkte können auch andere Frequenzen verwenden. Am Empfänger wird das Signal lediglich demoduliert, es sind keine weiteren Signalverbesserungsmaßnahmen erforderlich. Obwohl es nicht zwingend vorgeschrieben ist, können z. B. Techniken zur Rauschunterdrückung eingesetzt werden.

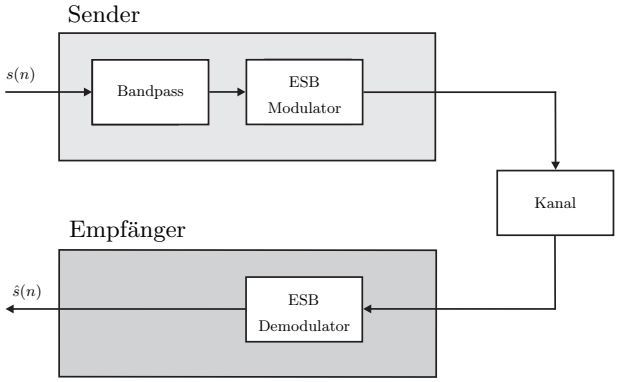

Abbildung 1: Übersicht des traditionellen Ansatzes.

Gemischter Ansatz

Das Konzept der gemischten analog-digitalen Übertragung weist sowohl Ähnlichkeiten mit der analogen als auch digitalen Unterwasser-Sprachübertragung auf. Das Konzept der gemischten Übertragungen für Audio und Sprache wurde bereits in [2] und [3] für höhere Datenraten und verschiedene Anwendungen untersucht. Für Unterwasseranwendungen wurde das Verfahren in [4] vorgestellt. Dem NATO-Standard gemeinsam ist die Übertragung eines analogen Sprachanteils mit Einseitenbandmodulation. Einige Sprachmerkmale werden jedoch wie bei digitalen Übertragungen mitgesendet. Die Verarbeitungsstruktur für das Simulationsmodell ist in **Abbildung 2** dargestellt. Im praktischen Einsatz müssen D/A- bzw. A/D-Wandler und entsprechende Schallwandler hinzugefügt werden. Die Idee ist, mittels LPC-Filterung die Hüllkurve aus dem Sprachsignal zu entfernen und das Restsignal analog zu übertragen. Das Restsignal wird zusätzlich einer Energieanpassung unterzogen, sodass es sowohl im Zeit- als auch im Frequenzbereich eine stark reduzierte Dynamik aufweist. Die direkte Übertragung der einzelnen LPC-Koeffizienten würde eine sehr große Datenrate erfordern, weshalb ein Codebuch verwendet wird, welches die Anzahl der möglichen Hüllkurven entsprechend der Codebuchgröße begrenzt. Um ein Codebuch zu erstellen, muss zuvor eine ausreichend große Datenbank mit mehreren Sprechern zur Verfügung stehen, um möglichst viele Hüllkurven zu generieren (z.B. die NTT Multilingual Speech Database 2002 [5]). Das Codebuch wird mit dem k-Means-Algorithmus [6] trainiert, und der LPC-Filter hat den Grad 8. Da die LPC-Koeffizienten jedoch empfindlich auf

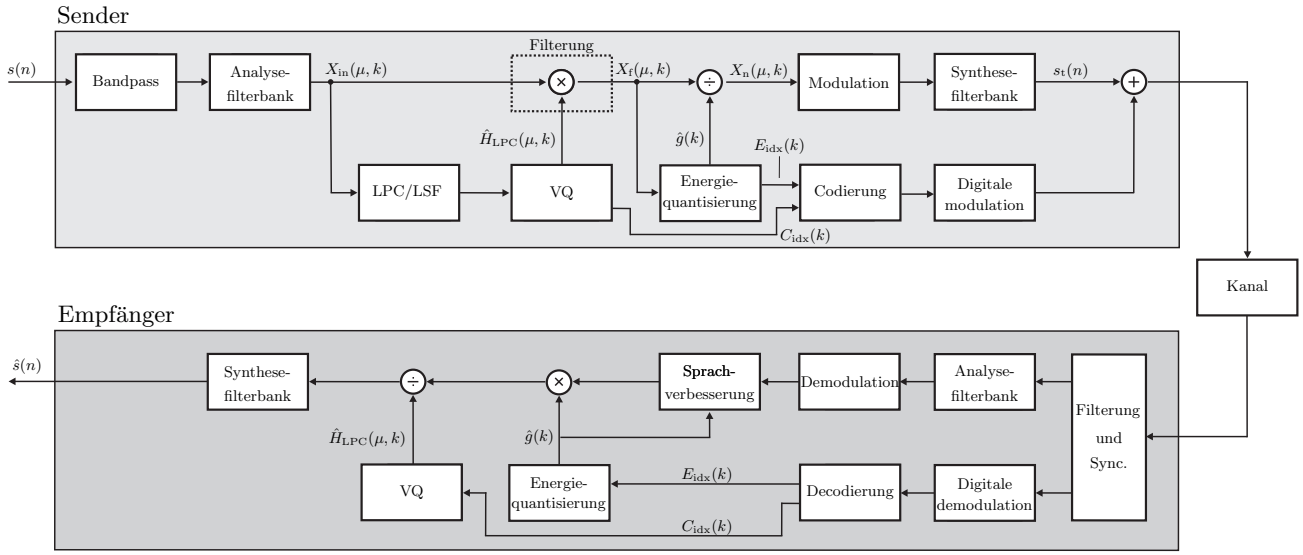

Abbildung 2: Übersicht über das gemischt analog-digitale Kommunikationssystem.

Quantisierung reagieren, werden im Codebuch anstelle von LPC *linear spectral frequencies* (LSF) gespeichert. Diese Formen sind äquivalent und können direkt ineinander umgewandelt werden, aber sie gewährleisten auch in vergleichsweise kleinen Codebüchern eine gute Leistung, da sie weniger anfällig für Quantisierungsrauschen sind. Das geweißte, energienormalisierte analoge Restsignal des gemischt analog-digitalen Ansatzes ist in **Abbildung 3** als Zeitsignal und Spektrum gezeigt.

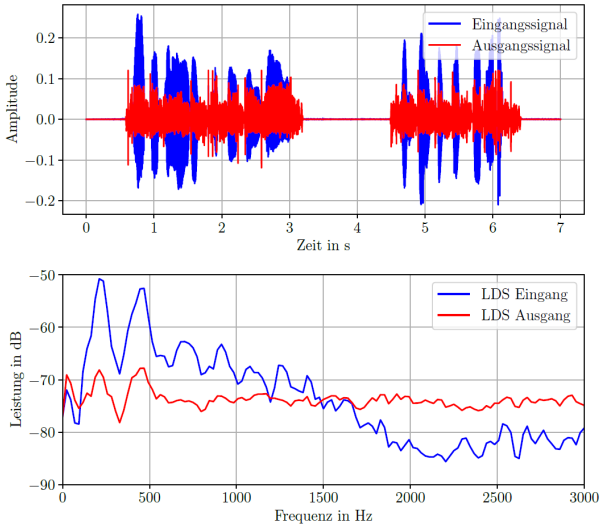

Abbildung 3: Sendesignal als Zeitsignal und Spektrum.

Für jeden Rahmen von Eingangssignalen wird die Sendesignalverarbeitung durchgeführt, sodass für jeden dieser Rahmen ein Codebuchindex und ein Energieindex extrahiert wird. Diese Indizes werden im digitalen Teil codiert, moduliert und zusammen mit dem modulierten analogen Restsignal über den Kanal zum Empfänger geschickt. Am Empfänger kann mit den Indizes für Codebuch und Energie sowie dem Restsignal die Sendesignalverarbeitung rückgängig gemacht und das Signal $\hat{s}$ rekonstruiert werden.

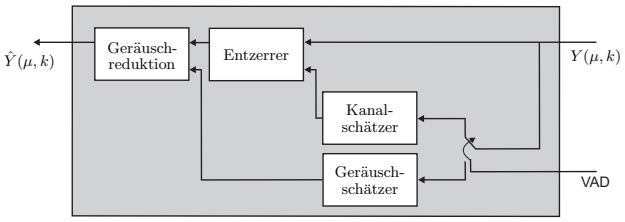

Abbildung 4: Übersicht des Sprachverbesserungsmoduls.

Sprachverbesserung

Verglichen mit dem traditionellen Ansatz können die Sprachverbesserungsmöglichkeiten auf der Empfangsseite deutlich erweitert werden. Durch die zusätzliche Übertragung der am Sender anliegenden Energie des Sprachsignals, welche für die Normalisierung des Restsignals genutzt wird, stehen am Empfänger Informationen über die senderseitige Sprachaktivitätsdetektion (*engl. Voice Activity Detection - VAD*) zur Verfügung. Eine sehr kleine Energie lässt darauf schließen, dass keine Sprache aktiv ist (VAD=0). Folglich werden in diesem Augenblick nur Hintergrundgeräusche empfangen. Ist die Energie jedoch hoch, befindet man sich in einer Phase mit aktiver Sprache (VAD=1) und wie aus **Abbildung 3** ersichtlich wird, ist hier die ungefähre Form des Sendesignals bekannt, was zu einer Entzerrmöglichkeit führt, da ein relativ flaches Spektrum zu erwarten ist.

Geräuschreduktion

Durch das Vorhandensein einer senderseitigen VAD vereinfacht sich die Schätzung des Hintergrundgeräuschpegels stark, da ein aufwendiges Trennen von Sprache und überlagertem Rauschen deutlich erleichtert wird. Der Schätzer für das Hintergrundgeräusch kann stark vereinfacht ausgeführt werden verglichen mit herkömmlichen Verfahren und zusätzlich relativ aggressiv agieren, da die Unterscheidung zwischen Sprache und Geräusch recht sicher getroffen werden kann, weil diese Information bereits vom Sender mitgeteilt wird. Die Schätzung des Geräuschspektrums aus dem

geglätteten Eingangsspektrum erfolgt je nach aktivem Sprachzustand durch:

$$\hat{N}(\mu,k)\Big|_{\mathrm{VAD}\,=\,0} = \tag{1}$$
$$\begin{cases} \hat{N}(\mu,k-1)\cdot\Delta_{\mathrm{inc}}, & \hat{N}(\mu,k-1) \le \overline{Ys(\mu,k)} \\ \hat{N}(\mu,k-1)\cdot\Delta_{\mathrm{dec}}, & \hat{N}(\mu,k-1) > \overline{Ys(\mu,k)}, \end{cases}$$

wobei $\overline{Ys(\mu,k)}$ das geglättete Eingangsspektrum bezeichnet. Δ_{inc} und Δ_{dec} sind Konstanten, welche die Geschwindigkeit der Änderung des Geräuschspektrums definieren. In Phasen aktiver Sprache wird die letzte Schätzung verwendet:

$$N(\mu,k)\Big|_{\mathrm{VAD}\,=\,1} = N(\mu,k-1). \tag{2}$$

Analog zu einer klassischen Geräuschreduktion wird mit dem so geschätzten Geräuschspektrum ein Wiener Filter [7] mit der Übertragungsfunktion $H_{\mathrm{n}}(\mu,k)$ betrieben

$$H_{\mathrm{n}}(\mu,k) = \max\left\{H_{\mathrm{min,n}},\, 1-\frac{\hat{S}_{bb}(\mu,k)}{\hat{S}_{yy}(\mu,k)}\right\}, \tag{3}$$

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

Entzerrung

Die spektrale Beschaffenheit des analogen Restsignals macht weiterhin eine begrenzte Entzerrung möglich. Das Restsignal ist durch die Filterung im Sender spektral ausgewogen, sodass davon ausgegangen werden kann, dass große Abweichungen der mittleren Leistung über alle Frequenzen durch den Kanal oder die Sende-/ bzw. Empfangshardware hervorgerufen werden. Da zum Filtern des Eingangsspektrums jedoch nur eine durch die Codebuchgröße eingeschränkte Anzahl an Vektoren zur Verfügung stehen, ist das Signal nicht vollständig weiß, sodass eine Schwelle definiert wird, ab wann ein Ausgleich der spektralen Leistung erfolgen muss. Im Gegensatz zur Geräuschreduktion wird eine Schätzung des Eingangssignalspektrums nur vorgenommen, wenn Sprache vorhanden ist. Das Eingangssignal wird zuerst zeitlich und dann über die Frequenzbins geglättet. Für die zeitliche Glättung gilt:

$$\begin{aligned}\overline{Y_{\mathrm{eq,t}}(\mu,k)} = {} & \Delta_{\mathrm{eq,t}}\cdot\overline{Y_{\mathrm{eq,t}}(\mu,k-1)} \\ & + (1-\Delta_{\mathrm{eq,t}})\cdot|Y_{\mathrm{in}}(\mu,k)|.\end{aligned} \tag{4}$$

Der Glättungsfaktor $\Delta_{\mathrm{eq,t}}$ wird in dB/s angegeben und so gewählt, dass nur eine relativ langsame Änderung zugelassen wird. Auf diese Weise kann erreicht werden, dass kurzfristige Schwankungen nicht zu stark ins Gewicht fallen. Zusätzlich zur zeitlichen Glättung, wird das Spektrum auch entlang der Frequenzachse geglättet. Der Glättungsfaktor für die Frequenzachse $\Delta_{\mathrm{eq,f}}$ beschreibt die zulässige Änderung von einem Frequenzbin zum nächsten. Das resultierende doppelt geglättete Spektrum wird mit $\overline{\overline{Y_{\mathrm{eq,f}}(\mu,k)}}$ bezeichnet. Von dem doppelt geglätteten Spektrum wird der Mittelwert $\overline{\overline{Y_{\mathrm{eq,mean}}(k)}}$ gebildet:

$$\overline{\overline{Y_{\mathrm{eq,mean}}(k)}} = \frac{1}{N}\sum_{\mu=0}^{N-1}\overline{\overline{Y_{\mathrm{eq,f}}(\mu,k)}}. \tag{5}$$

Dieser Mittelwert wird als Grundlage der Entzerrung genommen und daraus die Koeffizienten des Entzerrers bestimmt:

$$H_{\mathrm{eq}}(\mu,k) = \tag{6}$$
$$\begin{cases} H_{\mathrm{min,eq}} & \text{, wenn } \frac{\overline{\overline{Y_{\mathrm{eq,mean}}(k)}}}{\overline{\overline{Y_{\mathrm{eq,f}}(\mu,k)}}} < H_{\mathrm{min,eq}}, \\ H_{\mathrm{max,eq}} & \text{, wenn } \frac{\overline{\overline{Y_{\mathrm{eq,mean}}(k)}}}{\overline{\overline{Y_{\mathrm{eq,f}}(\mu,k)}}} > H_{\mathrm{max,eq}}, \\ \frac{\overline{\overline{Y_{\mathrm{eq,mean}}(k)}}}{\overline{\overline{Y_{\mathrm{eq,f}}(\mu,k)}}} & \text{, sonst.} \end{cases}$$

Zusätzlich werden noch $H_{\mathrm{min,eq}}$ und $H_{\mathrm{max,eq}}$ eingeführt. Diese Parameter bilden das Minimum und Maximum des Entzerrers. Das Minimum wird genutzt, weil das ankommende Restsignal nicht vollständig weiß ist. Wäre dieser Parameter nicht gesetzt, würden schon kleine Verzerrungen korrigiert und das Signal möglicherweise sogar verfälscht werden, da diese kleinen Abweichungen vom Mittelwert durchaus durch die nicht ideale Filteroperation zustande kommen können. $H_{\mathrm{max,eq}}$ ist ein Sicherungsparameter, ähnlich dem Geräuschteppich beim Wiener Filter, der die maximale auszugleichende Verzerrung definiert. In den Simulationen wurden diese Werte auf [-20 dB, 20 dB] festgelegt.

Simulation

Zur simulativen Evaluierung der beiden Sprachverbesserungen wurden zwei unterschiedliche Szenarien herangezogen. Zum Test der Geräuschreduktion wurde das Sendesignal einerseits mit weißem Rauschen versetzt und andererseits ein sinusförmiges Störsignal mit einer Frequenz von 1 kHz auf das Signal angewandt. In **Abbildung 5** ist zu erkennen, dass die Geräuschreduktion die Störung sehr stark dämpft. Durch die inverse Filteroperation im Empfänger ist das Übertragungsverfahren relativ tolerant gegenüber Rauschen, jedoch sind einzelne Geräuschkomponenten für den Teilnehmer auf der Empfangsseite sehr störend. Dies liegt darin begründet, dass nicht nur die inverse Filteroperation durchgeführt wird, sondern auch die Energie eines Sprachrahmens wieder angepasst wird. Dadurch werden jedoch auch additive Störungen mit angehoben oder abgesenkt im zeitlichen Verlauf des Signals. Aus diesem Grund ist die Geräuschreduktion an dieser Stelle von großer Wichtigkeit für den subjektiven Höreindruck. In **Abbildung 6**

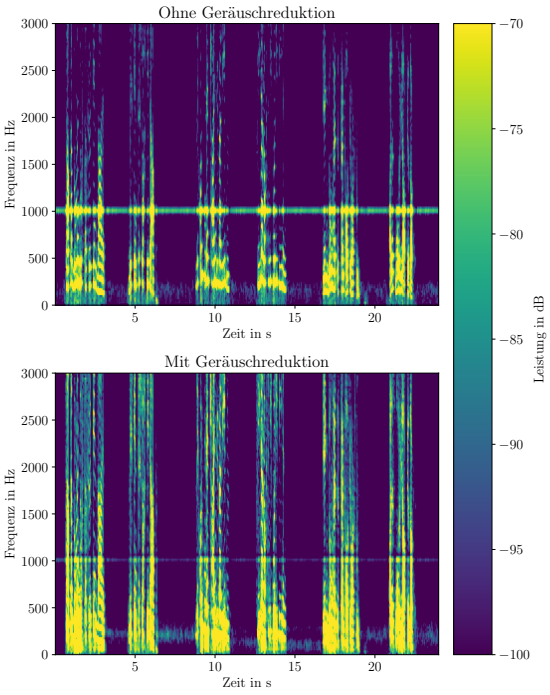

Abbildung 5: Empfangssignal mit und ohne Geräuschreduktion.

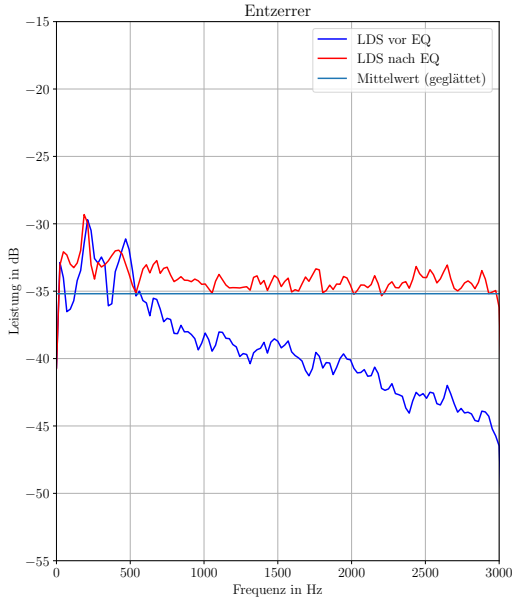

Abbildung 6: Leistungsdichtespektren vor und nach dem Entzerrer.

ist die zweite Simulation zu sehen. In dieser wurde simuliert, dass das Sendesignal beispielsweise durch Kanaleinflüsse oder die Sende-/Empfangswandler über die Übertragungsbandbreite um -10 dB abgesenkt wurde (dargestellt durch die blaue Kurve). Die Entzerrerstruktur ist in der Lage diese Verzerrung auszugleichen und so eine originalgetreuere Rekonstruktion am Empfänger zu ermöglichen (dargestellt durch die rote Kurve).

Zusammenfassung und Ausblick

In diesem Beitrag wurde gezeigt, dass, basierend auf dem gemischt analog-digitalen Übertragungsverfahren, eine begrenzte Entzerrung und Geräuschreduktion auf der Empfängerseite relativ einfach zu realisieren ist. Das energie- und frequenznormalisierte Signal wird im Kanal Störungen unterworfen, welche durch die Sprachverbesserungsalgorithmen gemindert werden. Der mitgesendete Energieindex zur Leistungsnormalisierung wird genutzt, um eine Sprachaktivitätsentscheidung am Empfänger zu treffen und somit je nach Zustand Kanal- oder Geräuschschätzer zu betreiben. Da sowohl der digitale, als auch der analoge Teil für die Rekonstruktion des Ursprungssignals notwendig sind, ist das Verfahren von einer zuverlässigen Übertragung der digitalen Informationen abhängig. Aus diesem Grunde ist eine Stabilisierung des digitalen Links ein wichtiges Ziel. Weiterhin ist der Aufbau eines realistischen Vergleichsszenarios zwischen dem Standardansatz und gemischten Ansatz angedacht. Dieses Vergleichsszenario sollte mit echten Daten und nicht auf Simulationen beruhen, sodass das Verfahren unter realistischen Bedingungen evaluiert werden kann. Anschließend sollte die Auswertung der so gewonnenen Daten durch Hörtests untermauert werden. Ebenso sollte zukünftig die genutzte Bandbreite des Verfahrens reduziert werden, da durch das hybride Verfahren sowohl digitale als auch analoge Daten parallel in unterschiedlichen Frequenzbändern übertragen werden müssen. Der analoge Teil nimmt dabei die gleiche Bandbreite in Anspruch wie im STANAG-Standard und die Bandbreite des digitalen Teils kommt dann noch dazu. Die spektrale Ausgewogenheit des Restsignals bietet jedoch Ansätze für Bandbreitenerweiterungsalgorithmen, sodass möglicherweise nicht mehr das vollständige Restsignal übertragen werden muss.

Literatur

[1] NATO Standardization Office: STANAG 1475 - Material interoperability requirements for submarine escape and rescue (2014)

[2] Hoelper, C. und Vary, P.: A New Modulation Concept for Mixed Pseudo Analogue-Digital Speech and Audio Transmissioin, ICASSP (2007)

[3] Rüngeler, Matthias: Hybrid Digital-Analog Transmission Systems: Design and Evaluation, ABDN (2015)

[4] Wisch, O. und Schmidt, G.: Mixed Analog-digital Speech Communication for Underwater Applications, 14th ITG Conference (2021)

[5] Multilingual Speech Database 2002, `https://www.ntt-at.com/product/speech2002/`, Abgerufen 03.05.2021

[6] Lloyd, S.: Least squares quantization in PCM, IEEE Transactions on Information Theory (1982)

[7] Y. Huang, J. Benesty, M. M. Sondhi (Eds.): *Springer Handbook of Speech Processing,* Springer Verlag (2008)