

Verbesserung der Sprachverständlichkeit bei analoger Unterwasserkommunikation durch Nutzung von digitalen Anfangs- und Endmarken

Ivor Nissen und Eric Kuhnt-Matthé

Bundeswehr Technical Centre for Ships and Naval Weapons, Maritime Technology and Research (WTD 71)

Email: IvorNissen@Bundeswehr.org, in@is.informatik.uni-kiel.de

In diesem Beitrag wird die Verwendung von digitalen Markierungen in Form von Prä- und Postambeln bei der analogen akustischen Sprachübertragung im Unterwasserbereich eingeführt, ähnlich der Quindar-Töne der NASA. Diese Erweiterung gestattet es, 80 Jahre nach den ersten analogen Sprachübertragungen in der Kieler Förde, den Schallkanal zu vermessen, Verzerrungen zurückzurechnen, um damit die Sprachverständlichkeit für den Hörer zu verbessern, aber auch durch die zeitliche Begrenzung mit Anrufbeantworter- und Verschleierungsfunktionen ein Unterwassertelefon aufzurüsten. Messungen in See wurden 2014 auch im Tiefwasser durchgeführt, Auswertungen werden vorgestellt.

Einführung in die Problematik

Analoger Unterwasserkommunikation kann seit Jahren kein Fähigkeitsaufwuchs bescheinigt werden. Diese Form der Nachrichtenübermittlung mittels menschlicher Sprache wird durch STANAG 1475 (alt: STANAG 1074) vereinheitlicht; die Modulationsart hat sich dabei seit der Anfangszeit in den 1930ern nicht verändert [DAGA17]. Eine nicht verstandene oder unbemerkte Sprachübermittlung kann nicht erneut gehört werden, es fehlt an einem Anfang und einem Ende, um eine automatisierte Abspeicherung zu gewährleisten. Einziges Hilfsmittel ist die Retransmission, ein unbefriedigender Zustand für die Sicherheit bei Tauchvorhaben.

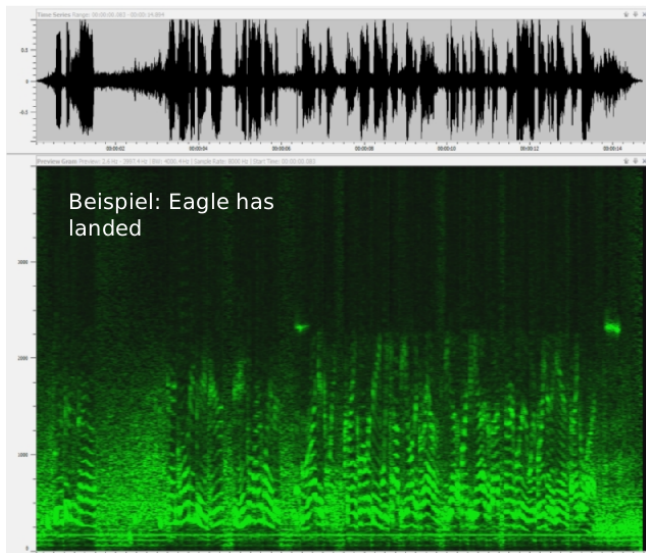


Abbildung 1: Der Quindar-Ton (CW) dient der Inband-Emulation der Push-to-talk-Taste (PTT) im Datenstrom. Hier am Beispiel „Eagle has landed“ (www.nasa.gov/connect/sounds/). Oberhalb des Sprachbandes sind im Spektrum die beiden Töne (s. Bildmitte) zu erkennen.

Ein Verbesserungsschritt wäre die Erweiterung der analogen Sprache vor und nach der Transmission um

digitale breitbandige Anfangs- und Endmarken im Übertragungs-Frequenzband, im weiteren als Prä- und Postambeln bezeichnet. Diese Möglichkeit der Inband-Markierung ist nicht neu. So benutzte die NASA für die Apollo-Missionen sogenannte Quindar-Töne, die der Emulation der Push-to-talk-Taste (PTT-Taste) in dem Datenstrom dienen, um eine Beschränkung auf ein Frequenzband und damit einen Übertragungsweg zu erzielen, wie es in Abbildung 1 ersichtlich ist.

Durch die Nutzung von digitalen Anfangs- und Endmarken generiert man mindestens die Möglichkeiten

- der Verbesserung der Sprachverständlichkeit durch Entzerrung,
- des maschinellen Ausschneidens zur Archivierung („Anrufbeantworter“),
- zur Verschleierung durch variierendes Zerhacken und
- der maschinellen Spracherkennung (Warnung, Dokumentation, Weiterverarbeitung)

Diese Marken sollten sehr kurz und nicht periodisch - also transient - sein. Dieser Beitrag dient der Weiterentwicklung des Unterwassertelephons (UT).

Nutzung von OFDM-Hadamard-Marken

In diesem Beitrag sollen kurze Bursts [SHA15] in Form von Inband-Hadamard-BPSK-Mehrträgersymbolen als Beispiel für transiente Marken genutzt werden, die Metadaten übertragen und neben der Sprache dekodiert werden können. Dabei kann die Anfangsmarke (Präambel) gleich der Endmarke (Postambel), oder wie hier genutzt Präambel ungleich der Postambel sein; wie in Abbildung 2 dargestellt. Es wird folgender Inhalt mittransportiert:

- Die Zeitdauer der Sprach-Transmission (zur Dopplerkompensation),
- eine Authentifizierung / Schlüsselnummer,
- Zeitstempel, Netzwerk-Adressen, Sender-Position,
- Kanalvermessungshilfen (Pilotöne).

Es ergeben sich damit folgende beiden Algorithmen:

Algorithmus Senden:

- Wähle Dimension, z.B. $N = 256$ und generiere Hadamard-Matrix für Pilotenwahl
- es gibt damit mind. N Spalten und damit mind. N Möglichkeiten um Authentifizierung, Adressen usw. abzubilden.
- Wähle hieraus zwei Spalten aus, Prä- und Post-

- spalte. (Nutze z.B. jede zweite Frequenzzelle zum Übertragen der digitalen Inhalte, Piloten)
- generiere BPSK-OFDM-Symbol für beide Marken.
- füge diese am Anfang und Ende der Sprache hinzu.

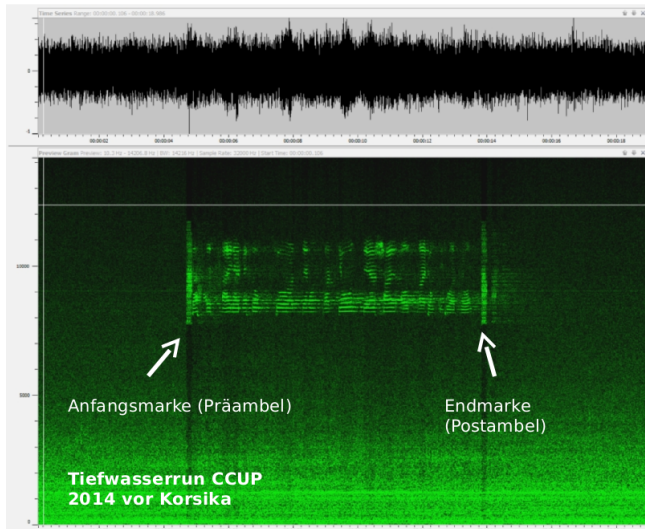


Abbildung 2: Hinzufügen digitaler Anfangs- und Endmarken im Übertragungs-Frequenzband zur analogen Sprache, hier am Beispiel der Tiefwasserversuche der WTD 71 im Sommer 2014 (Fahrt CCUP14).

Analog für das Empfangen und Auswerten:

Algorithmus Empfangen:

- Grob-Dopplerkorrektur durch f_{min} und f_{max}
- Dekodiere die zugehörigen digitalen Inhalte (Authentifizierung, Netzwerkadressen, Zeitstempel, Länge) als Metadaten
- Feindopplerkorrektur durch Längenparameter
- Bestimme $N = 256$ Kanalkoeffizienten bzgl. Frequenzbereich
- Generiere zugehöriges Sendesignal gleichen Inhalts durch Modulator; nutze dieses Replica zur Korrelierung als dritten Synchronisationsschritt zur
- Bestimmung des Sprachsegmentes (Ausschneiden)
- Speichere Sprachnachricht als Empfangs-Zeit-signal
- Ordne der abgespeicherten Sprachnachricht Metadaten sowie Empfangszeit zu
- Linearkombination aus Prä- und Post-Kanalwerten durch Nutzung der Piloten (Entzerrung)
- Korrigiere Sprachsignal anhand der Kanalkoeffizienten
- Speichere entzerrte Sprachnachricht als Aufbereitung ab
- Führe maschinelle Spracherkennung durch
- Dokumentiere die eingekommene Nachricht mit Inhalt und Verknüpfung zur Sprachnachricht, sowie der aufbereiteten Nachricht

Der Burst trägt damit Metadaten und Piloten zur Kanalvermessung gleichermaßen. Eine Kompatibilität zu Altgeräten ist gegeben. Eine Kommunikation mit einem Empfänger ohne digitalen Dekodern erfüllt die STANAG 1475, die Marken werden als Rauschklicks

am Anfang und Ende der Sprachtransmission durch den Zuhörer aufgefaßt. Fehlende digitale Marken führen ebenfalls zu der klassischen Form der STANAG 1475. Die Einzelschritte werden nun gesondert beschrieben:

Verbesserung der Sprachverständlichkeit durch Entzerrung beim Empfänger

Auf der Empfängerseite werden beide digitalen Marken, z.B. wie in [SHA15] beschrieben, demoduliert und dekodiert und durch den eingebauten Sendezweig wieder kodiert und moduliert. Damit liegt auf der Seite des Empfängers das gleiche Wissen, wie das des Senders vor. Der Empfänger hat das identische Sendesignal der Marken. Die beiden Zeitsignale der Transienten können in einem inkohärenten Verfahren als Replika zur Korrelation genutzt werden, um die Impulsantwort sowie die samplegenauen Start- und Endpunkte der Sprache zu bestimmen. Im Falle eines kohärenten Verfahrenszugangs können zusätzlich die Pilotsignale zur komplexen Kanalschätzung und damit zur Bestimmung der Übertragungsfunktion herangezogen werden. In der folgenden Abbildung 3 wird der Dekodierungsschritt illustriert. Dabei sind die Verteilungsdiagramme der Marken mit ihren komplexen Abbildungskomponenten für die Bit-Zustände 0 und 1 realisiert. Diese Metadaten werden mit dem Empfangszeitstempel in dem Empfangsunterwassertelefon in Textform dokumentiert.

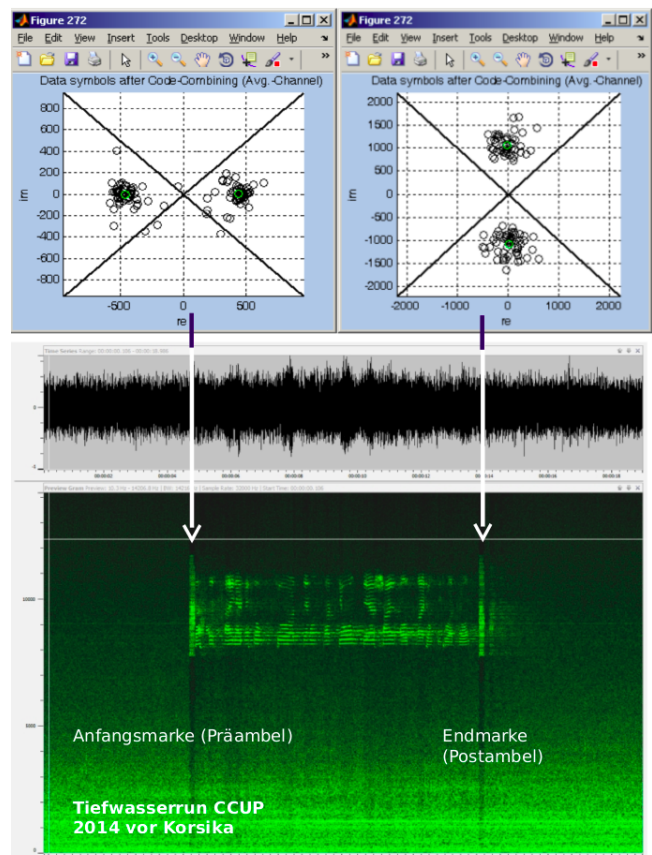


Abbildung 3: Inband-Hadamard-BPSK-Mehrträgersymbole ausgewertet, hier am Beispiel der CCUP14-Experimente. Dabei sind in diesem Fall die Phasen um 90 Grad gedreht, um festzustellen ob die Transmission beginnt oder endet.

Maschinelles Ausschneiden zur Archivierung

Durch die samplegenaue Bestimmung der Lage in den Zeitrohdaten sowie der Feststellung, ob die Marke den Anfang oder das Ende betrifft (beispielsweise durch den eingeführten 90 Grad-Phasenversatz), kann die Sprachnachricht aus dem Datenstrom exakt ausgeschnitten werden. Dabei wird mittels der übermittelten Längenmetainformation in der Präambel eine Dopplerkorrektur möglich. Mit der zugehörigen Entzerrung durch die errechnete Impulsantwort/Übertragungsfunktion kann damit die Originalsprachnachricht und die entzerrte Doppler-korrigierte Aufbereitungsnachricht als Audio-Sprachnachricht abgelegt werden.

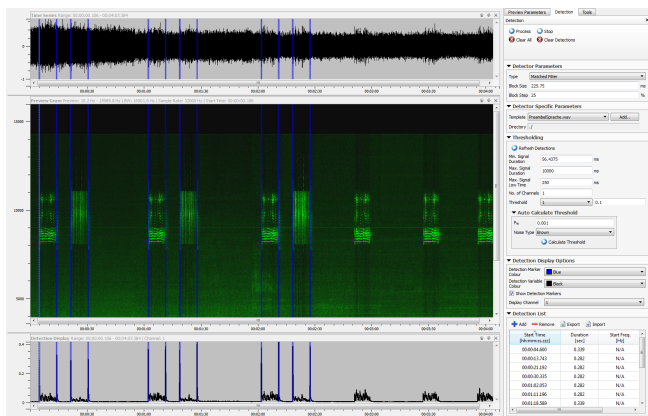


Abbildung 4: In den Meßdaten aus einem Fjordversuch im Sommer 2016 sind die ersten sechs Übertragungen, jeweils im Wechsel von offener und verschleierter Sprache mit Marken und die drei folgenden Sprachtransmissionen ohne Marken versehen. Mit dem dargestellten Empfangs-Algorithmus kann eine Segmentierung exakt vorgenommen werden.

Der Korrelationspeak ist durch die Breitbandigkeit der transienten Marken signifikant, in der Beispielsmessung wird ein Faktor 4 erzielt, siehe Abbildung 5.

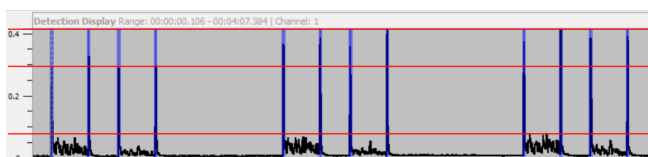


Abbildung 5: Eine Vergrößerung der Zeitkorrelationsachse zeigt die guten Korrelationseigenschaften der digitalen, breitbandigen Burst-Marken. Die Sprache korreliert in diesem Beispiel maximal bis 0.075, die Markenkorrelationswerte liegen zwischen 0.3 und 0.405.

Damit kann das beim Anrufbeantworter genutzte Abheben und Auflegen des Telefonhörers emuliert werden, in beiden Fällen hat man eine Begrenzung des Datenstroms auf ein festes Zeitintervall.

Verschleierung durch dynamisches Zerhacken

Durch die hochgenaue Ausschneidung kann nun auch eine Verschleierung durch Zerhackung durchgeführt

werden. Anders als die starren Zerhacker der frühen analogen Mobilfunkgenerationen, bietet sich durch die eingebettete digitale Mitsendung von Authentifizierungs-respektive Schlüsselnummern eine variable Anpassung. Ein Aufklärer muss damit nicht nur den Algorithmus, sondern auch über den aktuellen Schlüssel verfügen. Die spektrale Energiesignatur ist in Abbildung 4 dargestellt.

Maschinelle Spracherkennung

Auf Seiten des Empfängers können, neben der Dokumentation der Nachricht durch Abspeicherung der Sprache, die Metadaten um den maschinellen Spracherkennungstext bereichert werden. Auf Seiten des Senders kann zudem die gleiche Funktionalität genutzt werden, um Sprache in Text umzuwandeln, um diese digital als Komprimierungselement im Wasser zu übertragen und auf der Empfängerseite nach der Dokumentation diese durch das Telefon vorlesen zu lassen.

In beiden Fällen benötigt man eine maschinelle Spracherkennung. Da im Wasser sich das NATO-Alphabet in Verbindung mit der MXP 1 (Multi-national submarine and anti-submarine exercise manual) und den sogenannten 3-letter-codes durchgesetzt haben, ist eine Ganzwort-Lehrmethode effizient umsetzbar. Für diesen Beitrag wurde die Python-Bibliothek pyAudioAnalysis verwendet [PLOS ONE], die eine Palette von Auswertunterstützungswerkzeugen bereithält. So stehen mit Energie-, Entropy-, Frequenzschwerpunkt- und Fluxdetektoren insgesamt 34 Features zur Verfügung. Für das Training und Clustern stehen direkt in der Bibliothek *k-Nearest-Neighbor*- sowie über Fremdeinbindungen *Support Vector Machines*-, *Random forests*-, *Extra trees*- und *Gradient boosting*-Methoden zur Verfügung.

Zusammenfassung

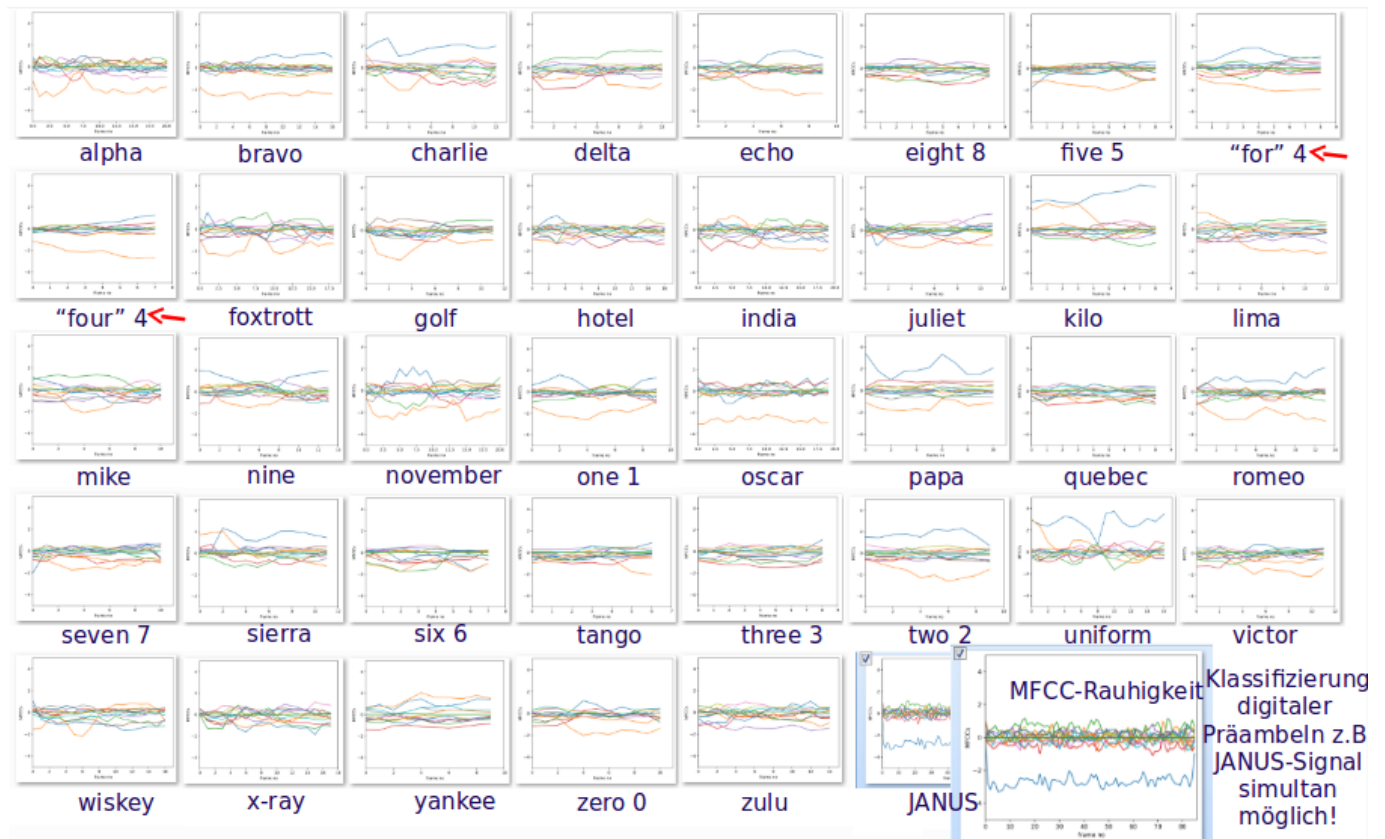
Führt man sich die ungenügende Sprachverständlichkeit vor Augen, die derzeit im horizontalen mehrwege- und damit echoversetzten Unterwasserschallkanal vorhanden ist, sowie die bei Tauchern, die Mundstücke zur Beatmung zwingend einsetzen müssen, so besteht akuter Handlungsbedarf einer Verbesserung in heutigen Unterwassertelefonen.

Die in diesem Beitrag vorgestellte Erweiterung der analogen Sprache um die Aussendung von digitalen Anfangs- und Endmarken kann hierzu beitragen. Vier Ansätze zur Entzerrung, dem maschinellen Ausschneiden für Archivierungszwecke, zur Verschleierung und der maschinellen Spracherkennung wurden realisiert und in See demonstriert. Damit ist der Stand der Technik erprobt und nun zugänglich. Zudem bleibt eine Kompatibilität zu Altgeräten bestehen. Durch die Mitsendung vom Sendezeitpunkt kann ferner auch die Entfernung geschätzt werden. Ein weiterer Zugang, der sich mit diesem Ansatz eröffnet, ist die analoge Sprach-Weiterleitung zur Überbrückung großer Distanzen. Die ausgeschnittene Sprachnachricht kann mit dem Rauschen verstärkt werden und dann mit neuen

digitalen Marken erneut in einem Relais, zum Beispiel einem Bodenknoten, ausgestrahlt werden.

Handlungsbedarf besteht weiterhin bei einer wortbasierten Segmentierung der Sprachnachricht als Fundament einer effizienten NATO-Alphabet-Erkennung sowie Sprachverständlichkeitsverbesserungen bei der mundstückbehafteten Taucherkommunikation, bei der die Marken unberührt bleiben. Prä- und Postamble nehmen nur die Effekte des Übertragungsschallkanals an und nicht die mundstückinduzierten Verstümmelungen der Sprache. Hierfür bedarf es neuer innovativer Zugänge.

Abbildung 6: Automatische Spracherkennung kann beispielsweise durch Extraktion von Mel-Frequency-Cepstrum-Koeffizienten, von der nicht frequenz-linearen Verarbeitung natürlicher Sprache im menschlichen Gehör abgeleitet werden. Hierzu werden die aus der analogen Unterwasserkommunikation extrahierten ersten 13 MFC-Features des NATO-Alphabetes dargestellt, die mittels der Python-Bibliothek *pyAudioAnalysis* und dem *k*-Nearest-Neighbor-Zugang erzeugt worden sind. Besonders an den Wörtern *FOR* und *FOUR* aber auch bei der *JANUS*-Transmission wird deutlich, dass diese Koeffizienten für eine maschinelle Warnung genutzt werden können und hierfür sinnvoll sind.



Literatur

[STANAG 1475] ED.1. *MATERIAL INTEROPERABILITY REQUIREMENTS FOR SUBMARINE. ESCAPE AND RESCUE* - ANEP/MNEP - 85

[DAGA17] Ivor Nissen. *Gertrude, 80 years underwater telephony*, DAGA, Kiel, März 2017

[SHA15] Ivor Nissen, *BURST COMMUNICATION - A SOLUTION FOR THE UNDERWATER INFORMATION MANAGEMENT*, HYDROACOUSTICS, Annual Journal, Vol. 18, Polish Acoustical Society, Gdansk Department, ISSN 1642-1817, 2015

[PLOS ONE] Theodoros Giannakopoulos, *pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis*. PLoS ONE 10(12): e0144610. doi:10.1371/journal.pone.0144610, 2015