



Automatische Transkription von UKW-Seefunkkommunikation zur Einsatzkoordination (ARTUS)

Teilvorhaben der DGzRS:
Spracherkennung und Transkription randomisierter Funkkommunikation
Schlussbericht: Teil II – Eingehende Darstellung

Thomas Lübcke & Oda Schliebusch-Jacob

Projektakronym	ARTUS
Titel des Verbundvorhabens	Automatische Transkription von UKW-See- funkkommunikation zur Einsatzkoordination
FKZ	13N15018
Laufzeit des Vorhabens	01.03.2019 – 30.06.2022

Berichtszeitraum	01.03.2019 – 30.06.2022
Erstellungsdatum	17.02.2023

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 13N15018 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Informationen des Zuwendungsempfängers



Deutsche Gesellschaft zur Rettung Schiffbrüchiger – DGzRS

Thomas Lübcke (Verbundkoordinator)

Werderstraße 2

28199 Bremen

Tel.: 0421-53 707-451

E-Mail: luebcke@seenotretter.de

Inhaltsverzeichnis

Abbildungsverzeichnis.....	IV
Abkürzungsverzeichnis.....	V
1. Überblick.....	1
1.1. Aufgabenstellung	1
1.2. Ausgangsbedingungen und Ablauf des Forschungsvorhabens	2
2. Eingehende Darstellung	4
2.1. Arbeitspakete und Forschungsergebnisse.....	4
2.1.1. AP 1 Systementwurf und Einsatzszenarien	4
2.1.2. AP 2 Erfassungssystem	9
2.1.3. AP 3 Spracherkennung und Programmierung	15
2.1.4. AP 4 Erstellung von Demonstratoren	23
2.1.5. AP 5 Implementierung und Erprobung im Einsatzkontext	27
2.1.6. AP 6 Flankierende F & E -Aktivitäten	29
2.1.7. Zusammenfassung und Plan-/Ist Vergleich	31
2.2. Wichtige Positionen des zahlenmäßigen Nachweises	33
2.3. Notwendigkeit und Angemessenheit der geleisteten Arbeit	33
2.4. Nutzung der Ergebnisse und Fortschreibung des Verwertungsplans.....	34
2.5. Bekanntgewordene Ergebnisse Dritter	35
2.6. Geplante und realisierte Veröffentlichungen	38
3. Pressespiegel.....	39
4. Literaturverzeichnis	40

Abbildungsverzeichnis

Abbildung 1 Arbeitspakete und Ablauf der Projektarbeiten	2
Abbildung 2 Einsatzszenarien	5
Abbildung 3 Systemanforderungen.....	6
Abbildung 4 Benutzeroberfläche des Erfassungssystems (Entwurf)	10
Abbildung 5 Klassisches Spracherkennungssystem	11
Abbildung 6 Data Augmentation durch Verrauschung	12
Abbildung 7 Beispiel für ein Transkript (Routinekommunikation) in englischer Sprache	13
Abbildung 8 Beispiel für ein Transkript (Routinekommunikation) in deutscher Sprache	14
Abbildung 9 Beispiel für ein gemischtsprachiges Transkript (DE-mixed).....	14
Abbildung 10 Aufteilung der ARTUS-Daten in Trainings- und Testdaten	16
Abbildung 11 Beispiel für Fehler im Vergleich zum Referenztranskript	17
Abbildung 12 Beispiel für quasi-redundante Textfolgen (Standardsicherheitsmeldung)	17
Abbildung 13 Beispiel für einsatzspezifische Textfolge	18
Abbildung 14 Entwicklung der WER in der ersten Projektphase	18
Abbildung 15 Uncased WER nach Technologiewechsel.....	21
Abbildung 16 ARTUS-Benutzeroberfläche (Borddemonstrator)	28
Abbildung 17 Plan-Ist Vergleich Zeitplanung	32

Abkürzungsverzeichnis

AHFE	<i>Applied Human Factors and Ergonomics Conference</i>
AI	<i>Artificial Intelligence</i>
AIS	<i>Automatic Identification System</i>
AMARIS	<i>Aeronautische und maritime Innovationsumgebung für interorganisationale Simulationen (BMBF-Forschungsprojekt)</i>
AP	<i>Arbeitspaket</i>
ARTUS	<i>Automatische Transkription von UKW-Seefunkkommunikation zur Einsatzkoordination</i>
ASR	<i>Automatic Speech Recognition</i>
ATC	<i>Air Traffic Control</i>
ATCO2	<i>Automatic Collection and Processing of Voice Data from Air-Traffic Communications (H2020-EU-Projekt)</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BOS	<i>Behörden und Organisationen mit Sicherheitsaufgaben</i>
BPOL	<i>Bundespolizei</i>
BRK	<i>Bayerisches Rotes Kreuz</i>
BRR	<i>Bremen Rescue Radio</i>
BSH	<i>Bundesamt für Seeschifffahrt und Hydrographie</i>
CML	<i>Fraunhofer-Center für Maritime Logistik und Dienstleistungen</i>
CSV	<i>Comma Separated Values</i>
CTC	<i>Connectionist Temporal Classification</i>
DFKI	<i>Deutsches Forschungszentrum für Künstliche Intelligenz</i>
DGzRS	<i>Deutsche Gesellschaft zur Rettung Schiffbrüchiger</i>
DLR	<i>Deutsches Zentrum für Luft- und Raumfahrt</i>
DLRG	<i>Deutsche Lebens-Rettungs-Gesellschaft e. V.</i>
DSC	<i>Digital Selective Call</i>
ECDIS	<i>Electronic Chart Display and Information System</i>
ELW	<i>Einsatzleitwagen</i>
F&E	<i>Forschung und Entwicklung</i>
FVT	<i>Fachstelle der WSV für Verkehrstechniken</i>
GPS	<i>Global Positioning System</i>
HAAWAI	<i>Highly Automated Air-Traffic Controller Workstations with Artificial Intelligence Integration (H2020-EU-Projekt)</i>
IAIS	<i>Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme</i>
IMO	<i>International Maritime Organization</i>
ISIDE	<i>Innovazione per la Sicurezza di Mare (Interreg Projekt)</i>
IuK	<i>Information und Kommunikation</i>
KI	<i>Künstliche Intelligenz</i>
MALORCA	<i>Machine Learning of Speech Recognition for Controller Assistance (H2020 EU-Projekt)</i>
MMS	<i>Mensch-Maschine Schnittstelle</i>
MRCC	<i>Maritime Rescue Co-ordination Centre</i>
NLP	<i>Natural Language Processing</i>
NotAs	<i>Unterstützung der Notrufannahme durch KI-basierte Sprachverarbeitung (BMBF-Forschungsprojekt)</i>
OSC	<i>On Scene Co-Ordinator</i>
PTT	<i>Push to talk</i>
SAR	<i>Search and Rescue</i>
SMA	<i>Swedish Maritime Administration</i>
SMCP	<i>Standard Marine Communication Phrases</i>
SPELL	<i>Semantische Plattform zur Intelligenten Entscheidungs- und Einsatzunterstützung in Leitstellen und beim Lagemanagement (BMWK-Forschungsprojekt)</i>
SVG	<i>Scalable Vector Graphics</i>

TAP	<i>Teilarbeitspaket</i>
UKW	<i>Ultrakurzwelle</i>
WER	<i>Word Error Rate</i>
WSV	<i>Wasserstraßen- und Schifffahrtsverwaltung</i>

1. Überblick

1.1. Aufgabenstellung

Die Koordinierung und Durchführung von Such- und Rettungsmaßnahmen auf See erfordern jederzeit eine zuverlässige Kommunikation zwischen allen Beteiligten. Gerade zu Beginn eines Einsatzes gehen viele Funkmeldungen gleichzeitig ein, während parallel wichtige Entscheidungen zu treffen sind, die von hoher Bedeutung für den weiteren Verlauf der einzuleitenden Maßnahmen sind. Können eingehende Sprachmeldungen nur unvollständig empfangen werden, so kann dies zu verzögerten und in der Folge zu suboptimalen Entscheidungen führen.

Das ARTUS-Projektkonsortium unter der Leitung der Deutschen Gesellschaft zur Rettung Schiffbrüchiger (DGzRS), Bremen in Zusammenarbeit mit den beiden Verbundpartnern Fraunhofer-Center für Maritime Logistik und Dienstleistungen (CML), Hamburg und dem Industriepartner RHOTHETA Elektronik GmbH, Murnau sowie mit Unterstützung des Unterauftragnehmers Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS), Sankt Augustin hat sich das Ziel gesetzt, ein Assistenzsystem zu entwickeln, das maritime Einsatzkräfte beim Prozessieren und Aufbereiten einsatzrelevanter Daten unterstützen soll, indem eingehende Funkmeldungen automatisch transkribiert und potenziellen Absendern zugeordnet werden. Dadurch soll sichergestellt werden, dass Funksprüche auch dann verifiziert werden können, wenn sie durch akustische Beeinträchtigungen an Bord und/oder in Phasen hoher Arbeitsbelastung, nicht oder nicht vollständig zu verstehen sind. Außerdem sollte das zu entwickelnde System die Möglichkeit bieten, die Inhalte eingegangener Meldungen recherchieren und ergriffene Maßnahmen dokumentieren zu können.

Der Fokus des Teilvorhabens lag primär in der Entwicklung und Programmierung der Spracherkennungskomponente auf der Basis eines tiefen neuronalen Netzes. Dazu waren ausgehend von domänenspezifischem Audiomaterial zunächst Datensätze zu erstellen, die im weiteren Verlauf dem Training eines bereits beim Unterauftragnehmer vorhandenen, vortrainierten Spracherkenners dienen sollten. Die Trainingsdatensätze waren auf der Grundlage von Not-, Dringlichkeits- und Routinemeldungen zu erstellen, die bei Bremen Rescue Radio (BRR), der von der DGzRS betriebenen Küstenfunkstelle, vom internationalen UKW- Not- und Anrufkanal 16 aufgezeichnet und bereitgestellt wurden und nach zuvor zu entwickelnden Transkriptionsregeln manuell zu verschriftlichen waren.

Für den Einsatz in verschiedenen Anwendungskontexten (an Bord verschiedener Einsatzmittel, in einer Küsten-/Landfunkstelle oder im Ausbildungsbetrieb im Simulator) waren in enger Zusammenarbeit mit potenziellen Endnutzern, repräsentiert u.a. durch die assoziierten Projektpartner¹ die Anforderungen für drei verschiedene Demonstratorvarianten zu ermitteln.

Das trainierte Spracherkennungssystem war anschließend gemeinsam mit den parallel vom Industriepartner zu entwickelnden Ortungsalgorithmen in die ebenfalls vom Industriepartner vorzubereitende Hardware zu integrieren und im Einsatzkontext zu erproben. Um das System auch an Bord von Schiffen unabhängig von der Verfügbarkeit einer stabilen Internetverbindung nutzen zu können, sollte es vollständig offline nutzbar sein.

Ergänzende F&E-Aktivitäten zu verwertungsrelevanten Fragestellungen waren begleitend zu verfolgen.

¹ Bayerisches Rotes Kreuz (BRK) Wasserwacht, Bad Kissingen; Bundesamt für Seeschifffahrt und Hydrographie (BSH), Hamburg; Bundespolizei (BPOL) See, Rostock; Deutsche Lebens-Rettungs-Gesellschaft e.V. (DLRG), Bad Nenndorf; Fachstelle der Wasserstraßen- und Schifffahrtsverwaltung (WSV) des Bundes für Verkehrstechniken (FVT).

Schwerpunkte der Projektarbeiten im Teilvorhaben der DGzRS lagen in den Arbeitspaketen (AP) 1-3 und 6. In den Arbeitspaketen 4 und 5 wirkte die DGzRS vor allem koordinierend und gemeinsam mit dem Unterauftragnehmer unterstützend mit.

Abbildung 1 gibt einen Überblick über die im Rahmen der Arbeitspakete zu erledigenden Projektarbeiten.

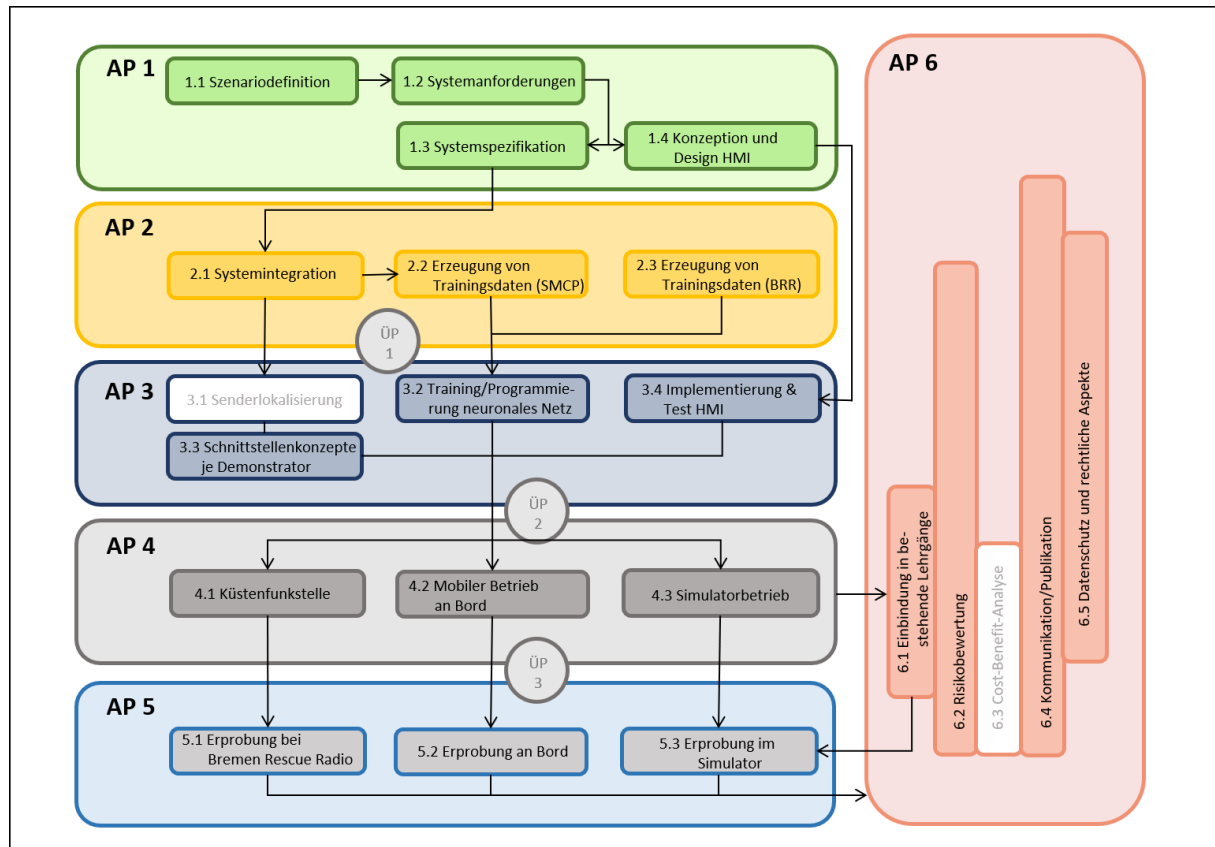


Abbildung 1 Arbeitspakete und Ablauf der Projektarbeiten

1.2. Ausgangsbedingungen und Ablauf des Forschungsvorhabens

Verglichen mit kommerziellen Spracherkennern erfordert der Betrieb eines solchen Systems im maritimen Kontext die Berücksichtigung domänenspezifischer Besonderheiten. Hier ist zunächst das spezielle Vokabular zu nennen, das von Standard-Spracherkennern nicht zuverlässig erkannt wird: Neben den von der International Maritime Organization (IMO) entwickelten Standardredewendungen (Standard Marine Communication Phrases, SMCP) sind eine Fülle an Eigennamen relevant, insbesondere Schiffsnamen und geographische Bezeichnungen. Weiterhin ist die Vielzahl an Akzenten bzw. Dialekten zu berücksichtigen: Die Verwendung der im internationalen Seeverkehr als „Lingua Franca“ geltenden englischen Sprache durch nicht Muttersprachler aus der ganzen Welt stellt besondere Anforderungen an die Spracherkennung. Im Einsatzgebiet der DGzRS ist weiterhin die häufig gemischte Verwendung von deutscher- und englischer Sprache sowie der Gebrauch platt- bzw. niederdeutscher Wendungen zu beachten. Beim Einsatz in der Binnenschifffahrt können weitere regionale Besonderheiten relevant werden.

Auch wenn der UKW-Seefunk im Vergleich zu anderen Kommunikationstechnologien sowohl aus Kostengründen sowie auch aufgrund besserer Verfügbarkeit nach wie vor das zentrale Kommunikationsmittel auf See darstellt, können sich die Entfernung zwischen Sender- und Empfänger, Witterungseinflüsse sowie Antennenhöhe einschränkend auf die Qualität des Audiosignals auswirken: Hier sind der z.T. hohe Grad der Verrauschung, Übersteuerungen

(„Clipping“) sowie Unterbrechungen eingehender Funkmeldungen zu nennen. Als weitere kritische Rahmenbedingungen für den Betrieb eines solchen Spracherkenners mit Senderlokalisierung sind schließlich zum einen die auf See bislang nicht flächendeckend verfügbare (Breitband-) Internetanbindung sowie die spezifischen Anforderungen an die zu verwendende Hardware relevant: Hier müssen Umwelteinflüsse wie Feuchtigkeit, Salzwasser und Seegang sowie das eingeschränkte Platzangebot an Bord berücksichtigt werden.

Das Gesamtvorhaben startete am 1. März 2019 und sollte gemäß Bewilligung bis zum 28. Februar 2022 abgeschlossen sein. Im Projektverlauf mussten in Folge der COVID-19-Pandemie einzelne Anpassungen der ursprünglichen Projektplanung vorgenommen werden. Als Rettungsdienst muss die DGzRS aufgrund ihrer hoheitlichen Aufgabe einem besonderen Schutzziel gerecht werden, so dass mit Beginn der COVID-19-Pandemie der Zugang sowohl an Bord der Seenotrettungskreuzer als auch an Land bei Bremen Rescue Radio und im Simulatorzentrum streng beschränkt werden musste. Obwohl sich die pandemieinduzierten Auswirkungen bereits in der ersten Projekthälfte bemerkbar machten, konnten alle Meilensteinziele erreicht werden.

Der Aufrechterhaltung des laufenden Betriebs in der Seenotleitung/ bei Bremen Rescue Radio sowie auf den Rettungskreuzern musste aus Infektionsschutzgründen absolute Priorität vor den Forschungsaktivitäten eingeräumt werden, so dass u.a. die Einbindung der Mitarbeiter von Bremen Rescue Radio in AP 3 nicht im geplanten Ausmaß erfolgen konnte. Ebenso konnten weder größere Übungen auf See noch im Simulator der DGzRS stattfinden und infolgedessen auch kein entsprechendes Funkdatenmaterial generiert werden. Die Schwierigkeiten bei der Generierung von Trainingsdaten für den Spracherkenner wurden durch die Einstellung einer weiteren Wissenschaftlichen Mitarbeiterin aufgefangen. Auf diese Weise konnte der Bestand an manuell transkribierten Daten zügig erhöht werden.

Durch die strengen Zugangsbeschränkungen zu den Räumlichkeiten an Land (Bremen Rescue Radio, Simulator)² und an Bord der Rettungskreuzer konnte die gesamte Installations- und Erprobungsphase der AP 4 und 5 nicht wie geplant ablaufen.

Auswirkungen der COVID-19-Pandemie machten sich zusätzlich auch bei der Beschaffung von Hardwarekomponenten für die Fertigung der Demonstratoren bemerkbar. Durch Lieferverzögerungen waren Rechnerkomponenten erst Anfang 2022 verfügbar, so dass die Demonstratoren nur durch die Gewährung einer ausgabenneutralen Projektverlängerung durch den Projektträger um 4 Monate bis zum 30. Juni 2022 im Projektzeitraum fertiggestellt werden konnten.

Aufgrund der verspäteten Fertigstellung der Demonstratoren und der weiterhin aufrechtzuerhaltenden Infektionsschutzmaßnahmen konnten die Demonstratoren - trotz Verlängerung - nicht im mehr im Projektzeitraum installiert und nicht im realen Einsatzkontext getestet werden. Eine Erprobung des ARTUS-Systems konnte daher nur unter Laborbedingungen erfolgen. Hierzu wurde beim Industriepartner ein Versuchsaufbau realisiert, in dem mit simulierten Datenströmen die Funktionalität der Demonstratoren erfolgreich getestet und potenziellen Endanwendern in Live-Onlinepräsentationen vorgestellt werden konnte.

In den folgenden Abschnitten werden die für das Teilvorhaben der DGzRS relevanten Teilarbeitspakete (TAP) hinsichtlich ihrer intendierten Zielsetzungen sowie der erreichten Arbeitsergebnisse vorgestellt. Den Abschluss hierzu bildet ein grafischer Plan-Ist-Vergleich der ursprünglichen Zeitplanung mit der tatsächlichen Projektrealisation.

An die Darstellung der Arbeitspakete schließen sich Ausführungen zu wichtigen Positionen des zahlenmäßigen Nachweises, zur Notwendigkeit und Angemessenheit der geleisteten

² Diese mussten nicht nur für Externe, sondern für auch abteilungsfremde Mitarbeiter der DGzRS fast bis zum Ende der Projektlaufzeit aufrechterhalten werden.

Arbeit, zur Nutzung der Ergebnisse und Fortschreibung des Verwertungsplans, zu bekanntgewordenen Ergebnissen Dritter sowie zu geplanten und realisierten Veröffentlichungen an.

2. Eingehende Darstellung

2.1. Arbeitspakete und Forschungsergebnisse

2.1.1. AP 1 Systementwurf und Einsatzszenarien

Gegenstand des ersten Arbeitspaketes waren die Erstellung und Dokumentation der Szenariodefinitionen und der daraus resultierenden Systemanforderungen, die ihrerseits die Grundlage für die Spezifikation der zu verwendenden technischen Komponenten bildeten. Hierbei galt es, die beiden wesentlichen Funktionen Spracherkennung/Transkription auf der einen Seite und Objektlokalisierung/Absenderermittlung auf der anderen Seite hard- und softwaretechnisch zu integrieren. Um das ARTUS-System für den praktischen Einsatzkontext anwenderfreundlich zu gestalten, war die Mensch-Maschine-Schnittstelle (MMS) unter Berücksichtigung der speziellen Bedürfnisse der potenziellen Endnutzerguppen zu konzipieren. Die intensive Einbindung der Endanwender aus den Bereichen Schiffsführer/Bootsführer, Leitstellenpersonal, Radio Operatoren, Simulator-Instruktoren und Trainer erfolgte u.a. durch Einladung zu einem gemeinsamen Workshop in Bremen. In entsprechenden Arbeitsgruppen wurden Anforderungen, Eigenschaften und gewünschter Funktionsumfang eines Assistenzsystems für den maritimen Such- und Rettungs- Dienst diskutiert und evaluiert.

Unter Beteiligung aller drei Verbundpartner inklusive des IAIS als Unterauftragnehmer wurden in AP 1 alle erforderlichen Arbeitsschritte gemeinsam konzipiert und die Grundlagen für die weiteren Arbeitspakete geschaffen. Die DGzRS war gemeinsam mit dem Unterauftragnehmer in alle vier Teilarbeitspakete involviert, die im Folgenden näher erläutert werden.

TAP 1.1 Szenariodefinition

In diesem Teilarbeitspaket sollten nutzungskontext-spezifische Szenarien ausgearbeitet werden, die als mögliche Anwendungsfälle die Basis für die Formulierung der Systemanforderungen bilden.

Um ein breites Spektrum möglicher Anwendungskontexte sicherzustellen, wurde auf die Definition eines alleinstehenden Zielszenarios verzichtet. Alternativ wurden drei Anwendungsfälle beschrieben, mit denen die Funktionsfähigkeit des ARTUS-Systems in zwei Entwicklungsstufen aufgezeigt werden kann:

- im Online-Betrieb für die Küstenfunkstelle und im SAR (Search and Rescue) -Simulator
- als integrierte Offline-Lösung für den Bordbetrieb

Den Ausgangspunkt für die weiteren Arbeiten und damit die Grundlage für die Formulierung der Systemanforderungen an die Demonstratoren bildeten damit drei nutzungskontext-spezifische Szenarien, die unter Einbeziehung der potenziellen Endanwender definiert wurden:

1. Szenario I: Demonstrator Küstenfunkstelle (BRR): „*Hilfloser Mitsegler*“
2. Szenario II: Demonstrator Bordbetrieb „*OSC-Lage*“
3. Szenario III: Demonstrator SAR *Simulator* „*Unterstützung der Übungsleitung bei komplexem Simulationsszenario*“

Szenario I Hilfloser Mitsegler <i>Demonstrator Küstenfunkstelle</i>	Szenario II: OSC-Lage <i>Demonstrator Bordbetrieb</i>	Szenario III: Unterstützung der Übungsleitung bei komplexem Simulationsszenario <i>Demonstrator Simulation (SAR Simulator)</i>
<p>Der Skipper einer Segelyacht fällt zwischen Kiel und Fehmarn über Bord. Dem noch an Bord befindlichen Mitsegler gelingt es jedoch nicht, den im Wasser treibenden Skipper wieder zurück an Bord zu holen. Da der Mitsegler weder mit der Bedienung des Bootes, noch den darauf befindlichen Rettungs- und Alarmierungsmitteln vertraut ist, greift dieser zum UKW-Sprechfunkgerät, drückt die PTT-Taste und ruft „Hallo? Hilfe! Hört mich jemand? SOS! Hallo?“ Diese Nachricht wird über den voreingestellten internationalen UKW-Not- und Anrufkanal 16 ausgesendet. Ein Radio Operator von Bremen Rescue empfängt die Notmeldung, antwortet und fragt strukturiert alle einsatzrelevanten Informationen von dem Mitsegler ab. Aufgrund mangelnder seemännischer Kenntnisse ist dieser jedoch nicht in der Lage, alle notwendigen Fakten mitzuteilen; so kann der Mitsegler unter anderem keine Angaben zur aktuellen Position des Schiffes machen. Erschwert wird die Kommunikation zwischen dem Havaristen und Bremen Rescue Radio durch Windgeräusche am Mikrofon des Havaristen sowie lautes Schlagen des Segels und hochfrequentes Schwingen der Fallen und Wanten.</p>	<p>Eine komplexe maritime Schadenslage als Einsatzleiter (OSC = On-Scene Co-ordinator) koordinieren zu müssen, gehört selbst für erfahrene Schiffsführer zu den herausfordernden Aufgaben; geht sie doch über die Organisation des eigenen Schiffsbetriebs weit hinaus und verlangt die Wahrnehmung einer mehrere Schiffe und Luftfahrzeuge umfassenden Führungsaufgabe. Neben der Initiierung zielführender Maßnahmen gehört vor allem die konsequente Lageerfassung zu den wesentlichen Aufgaben eines OSCs. Hierfür müssen häufig die Lageinformationen verschiedener Einsatzabschnitte eingeholt und verarbeitet werden. Ein vom OSC aufgestellter Kommunikationsplan trägt dazu bei, dass die einzelnen Einsatzabschnitte auf dem/den zugewiesenen UKW-Funkkanal/-kanälen ungestört operieren können. Der OSC kann wiederum einzelne Einsatzabschnittsleiter jederzeit gezielt über die individuell zugewiesenen Funkkanäle erreichen. Das Monitoring und die Dokumentation mehrerer parallelaufender UKW-Funkkonversationen bindet erhebliche personelle und/oder kognitive Ressourcen.</p>	<p>Das in den SAR-Simulator integrierte ARTUS-System wird die Kommunikation aller Cubicles über die Laufzeit eines Übungsszenarios von allen dort vorhandenen Sprechstellen (VoIP) automatisch transkribieren und eine Zuordnung zu den absendenden Cubicles resp. den geladenen Einheiten ermöglichen. Eine Suchfunktion, die sowohl die transkribierten Gesprächsinhalte als auch die beteiligten Sprechstellen umfasst, erlaubt es im laufenden Übungsbetrieb sowie während des Debriefings gezielt zu den einzelnen Phasen des Szenarios zu springen. Dies ermöglicht den Instruktoren jederzeit einen schnellen gedanklichen Wiedereintritt in die Gesamtlage, erleichtert die Moderation des Debriefings und stärkt die Nachvollziehbarkeit der einzelnen Übungsphasen. Über die SAR-Simulation hinaus kann das ARTUS-System die Lehrkräfte bei der Verwendung des Simulators in der Aus- und Weiterbildung unterstützen. So erlaubt die automatische Transkription beispielsweise eine objektivierte Prüfungsdokumentation bei SMCP-Lehrgängen bzw. beim Training allgemeiner maritimer Kommunikation oder kommunikationsintensiver Verfahren.</p>

Abbildung 2 Einsatzszenarien

Mit Szenario I wird abgebildet, wie das bei Bremen Rescue Radio zu installierende ARTUS-System auch bei widrigen akustischen Bedingungen in der Lage ist, alle Funkmeldungen eines Havaristen mittels Spracherkenner automatisch zu transkribieren. Über die Funktion der Senderlokalisierung ist es zudem möglich, die Position des Absenders der Meldung zu ermitteln. Dies ist insbesondere dann von Bedeutung, wenn ein Havarist, z.B. aufgrund mangelnder seemännischer Kenntnisse oder in einer Paniksituation nicht in der Lage ist, seine aktuelle Position korrekt zu übermitteln. Die Lokalisierung erfolgt anhand des UKW-Signals mittels Kreuzpeilung über im Seegebiet des Havaristen fest installierte Peilantennen. Die sofort alarmierten Rettungskreuzer können ohne suchbedingte Zeitverluste direkt zum Einsatzort fahren.

Szenario II beschreibt die Nutzung des ARTUS-Systems im Bordbetrieb. Ein typischer Anwendungsfall ist die sogenannte OSC (On Scene Co-ordinator) Lage. Bei einer komplexen maritimen Schadenslage hat der OSC als Einsatzleiter die Aufgabe, neben der Führung des eigenen Schiffes die Such- und Rettungsmaßnahmen in einem bestimmten Einsatzgebiet zu koordinieren. Die Regelung und Überwachung der Kommunikation mit und zwischen den beteiligten seegehenden und fliegenden Einheiten stellt dabei eine wesentliche Aufgabe dar und stellt hohe kapazitative Anforderungen an den OSC. Das ARTUS-System kann durch die automatische Transkription aller eingehenden Funkmeldungen und gleichzeitiger Absenderlokalisierung zu einer erheblichen Entlastung des Einsatzleiters beitragen.

Den dritten Anwendungsfall stellt schließlich die Integration des ARTUS-Systems in den von der DGzRS betriebenen SAR-Simulator dar. Während eines Simulator-Trainings müssen die Instruktoren einerseits jederzeit einen plausiblen Übungsverlauf sicherstellen und andererseits die gleichzeitig laufende Kommunikation in und zwischen den Übungsteilnehmern überwachen. Zwar werden die Übungen zum Zwecke eines fundierten Feedbacks aufgezeichnet. Diese Aufzeichnungen sind aber erst am Ende der Übung verfügbar und bieten keine simultane Dokumentation des Übungsgeschehens.

Die automatische Transkription der Kommunikation durch das ARTUS-System sowie die integrierte Zuordnungsmöglichkeit zu den Sprechern in den einzelnen Kabinen hingegen ermöglicht es den Instruktoren, zu jeder Zeit den Überblick über das gesamte Übungsgeschehen zu behalten und kurzfristig auf die Entwicklung des Trainings zu reagieren.

Über die Dokumentationsfunktion des ARTUS-Systems wird es darüber hinaus möglich sein, im anschließenden Debriefing mittels Filter bzw.- Suchtools schnell und gezielt auf einzelne Kommunikationssequenzen zugreifen zu können.

TAP 1.2 Identifikation von Systemanforderungen:

In diesem Teilarbeitspaket sollten auf der Grundlage der drei zuvor definierten Musterszenarien die Systemanforderungen für die jeweiligen Demonstratoren identifiziert werden.

Mit der Entwicklung verschiedener Demonstratorvarianten wurde den sehr unterschiedlichen Einsatzkontexten im Bereich des maritimen SAR-Dienstes und der Wasserrettung Rechnung getragen. Hierbei waren insbesondere die verschiedenen sprachlichen (englisch und deutsch im UKW-Seefunk vs. ausschließlich deutsche Sprachmitteilungen, ggf. in verschiedenen Dialekten im Binnenschiffahrtfunk) sowie die technischen und sonstigen Rahmenbedingungen zu berücksichtigen (Senderleistung und -reichweite, Sprachqualität, bauliche Eigenarten des Einsatzmittels).

Auf der Grundlage, der im vorangegangenen TAP identifizierten Anwendungsszenarien wurden in einem iterativen Prozess die Anforderungen an das ARTUS-System formuliert. Hierbei war zwischen allgemeinen Systemanforderungen an die Funktionalitäten des Gesamtsystems und abgeleiteten technischen Anforderungen an die drei unterschiedlichen Nutzungskontexte der zu entwickelnden Demonstratoren (an Bord, an Land, im SAR-Simulator) zu unterscheiden (vgl. Abb.3).

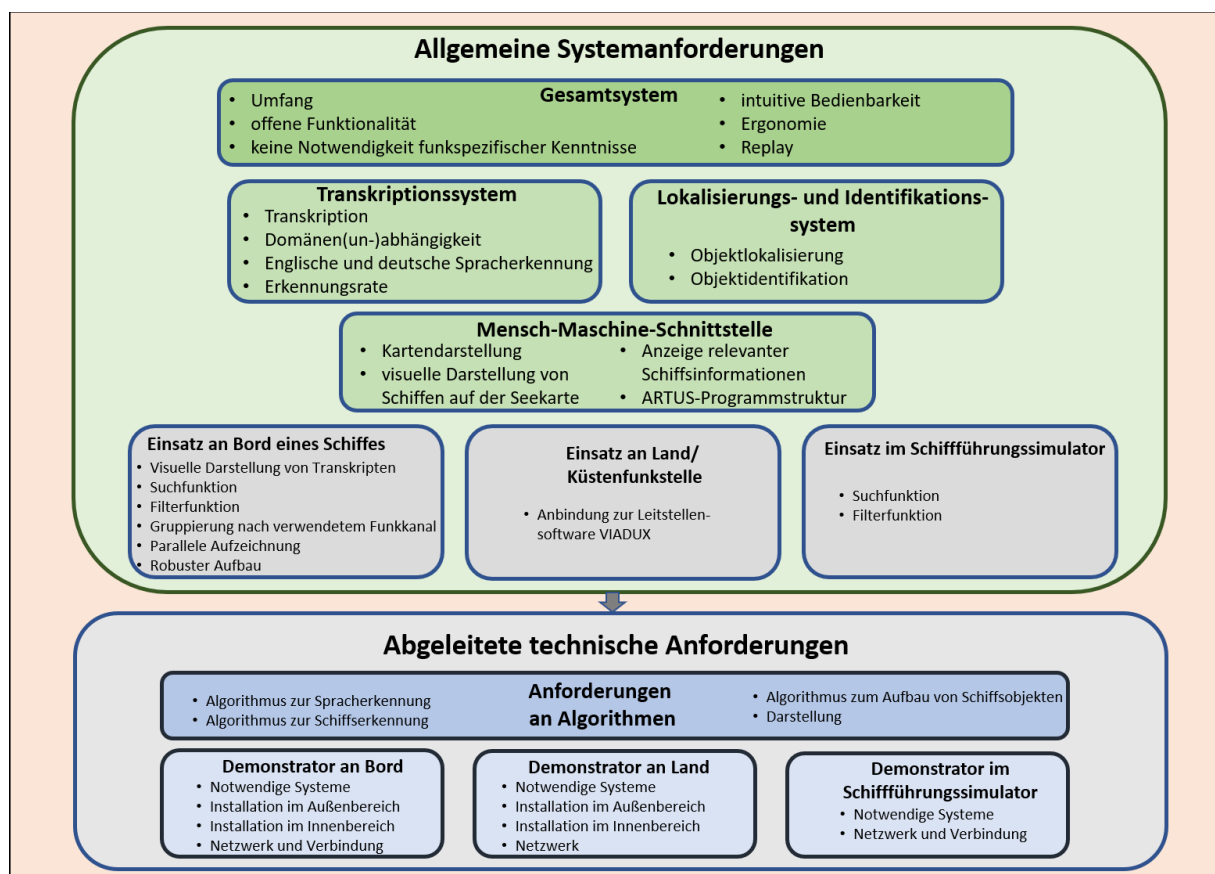


Abbildung 3 Systemanforderungen

Auf der Basis von „User Stories“ wurden die allgemeinen und kontextbezogenen Systemanforderungen formuliert. Hierbei wurden die potenziellen Endnutzer eng in den Prozess eingebunden: Im bereits erwähnten Workshop hatten sie die Gelegenheit, Anforderungen an das ARTUS-System unter der übergeordneten Leitfrage: „Wie sollen die von ARTUS generierten

Informationen aufbereitet und dargestellt werden, damit sich die Arbeitsbedingungen von Endanwendern im jeweiligen Nutzungskontext verbessern?“ zu diskutieren und zu evaluieren.

Folgende Funktionsbereiche wurden diskutiert:

- Transkriptanzeige (chronologische Sortierung/Zeitstempel, Priorisierung von Signalwörtern, Schriftgröße, Hervorhebungen, Edit-Funktion)
- Aufzeichnung Funksprüche (Aufzeichnung der Audiospur aller Funksprüche, Repeat-Funktion)
- Suchfunktion (Suche nach frei wählbaren Wörtern in der Transkriptionshistorie)
- Filterfunktion (Filterung Transkriptionshistorie nach definierten Begriffen)
- Übersetzungsfunktion (Übersetzung Englisch-> Deutsch)
- Visualisierung auf der Seekarte (ECDIS³, eigene Schiffsposition, Zoomfunktionen, Distanzeinblendung, Ausgrauen)
- Peilung/Lokalisierung (Peilstrahl, Erkennungswahrscheinlichkeit, Zusatzinformationen)
- Ergonomie (Tag/Nacht Modus, Auswahl der Hardware, Bedienelemente etc.)

Um die Erwartungen der Anwender zu fokussieren, wurden diese gebeten, entsprechende Anforderungen näher zu spezifizieren/zu ergänzen und diese anhand der Skala

- „must-have“
- „nice-to-have“
- „unwichtig“
- „störend“

zu bewerten.

Im Vergleich von Bord- und Landpersonal (hier: BRR und MRCC⁴) konnte eine relativ große Übereinstimmung bezüglich der „must-have“ Funktionen identifiziert werden. So wurden u.a. die folgenden Funktionalitäten von beiden Endnutzergruppen als zwingend erforderlich eingestuft:

- klare Trennung eingehender Funksprüche voneinander
- möglichst weit zurück reichende Transkriptionshistorie
- Darstellung bestimmter Zusatzinformationen (Uhrzeit, Peilung, Anrufkanal),
- Aufzeichnung der Audio-Spur aller Funksprüche inklusive Replay Funktion
- Darstellung lokalisierter Sender und Peilsysteme auf einer elektronischen Karte

Die größten Abweichungen ergaben sich im Bereich der Ergonomie: Während für das Bordpersonal aufgrund besonderer Einsatzbedingungen die Möglichkeit des Wechsels zwischen Tag-/Nachtmodus als „must-have“ gefordert wurde, ist dies für den Einsatz bei Bremen Rescue Radio erwartungsgemäß eher „unwichtig“. Gleiches gilt für bestimmte Bedienelemente: Hier wünscht sich das Bordpersonal zwingend eine Bedienung des ARTUS-Systems über Tasten, Steuerkreuz, Trackball oder ähnlichem, dagegen wird eine ausschließliche Steuerung mittels „Touch“ an Bord wegen der besonderen Einsatzbedingungen (z.B. Tragen von Handschuhen) sogar als „störend“ eingestuft, während dies für das Personal von Bremen Rescue Radio keine entscheidende Rolle spielt.

Im Ergebnis konnte der Forschungsverbund nun auf ein differenziertes Anforderungsprofil zurückgreifen, das im nächsten TAP in das Systemspezifikationsdokument zu überführen war.

³ = Electronic Chart Display and Information System

⁴ = Maritime Rescue and Co-ordination Centre

TAP 1.3 Systemspezifikation

In diesem TAP waren die zur Realisierung der in TAP 1.2 definierten Systemanforderungen erforderlichen technischen Komponenten zu definieren, d.h. es wurde festgelegt, auf welche Weise die Bereiche Spracherkennung/Transkription einerseits und Lokalisierung/Absenderermittlung andererseits in den zu entwickelnden Demonstratorvarianten integriert werden sollten. In diesem Zusammenhang war zunächst zu spezifizieren, welche externen Datenquellen über zu definierende Schnittstellen einbezogen werden, welche Geräte hierfür erforderlich sind und wie die Schnittstellen zwischen Hard- und Software auszugestalten sind. Die DGzRS und der Unterauftragnehmer wirkten hier unterstützend beim Entwurf bzw. der Erstellung der Spezifikationsdokumente mit.

Im ARTUS-System verarbeitet der Spracherkenner das eingehende analoge Funksignal, das er von einem Seefunkgerät, einem Funkpeiler oder als Audio-Datenstream von einem Seefunk-Relais empfängt und liefert als Output eine Nachricht in Textform sowie einen Sprachmuster-Code, der eine eindeutige Identifikation zugehörig zum Funkspruchmuster enthält. Als Datenquellen für die Objektlokalisierung dienen ein oder mehrere Funkpeiler. Das eingehende Signal repräsentiert neben weiteren Daten eine Peilung, Signalstärke und die eingestellte Frequenz. Der Anschluss eines Kompasses zur Ermittlung der eigenen Ausrichtung (Heading) ist möglich, so dass aus der relativen Peilung eine magnetisch nordbezogene Peilung errechnet werden kann. Als weitere Datenquellen dienen ein GPS-Empfänger zur Ermittlung der eigenen Position sowie ein AIS-Empfänger zur Positionsermittlung von (anderen) Schiffen. Die eingehenden Signale werden über verschiedene Algorithmen verarbeitet. Die Datenausgabe kann in einem Kartendisplay erfolgen und liefert u.a. die Ergebnisse zur Spracherkennung (transkribierter Funkspruch) und zur Objektlokalisierung (zur Identifizierung eines oder mehrerer Schiffe mit der jeweiligen Erkennungswahrscheinlichkeit, Absender des aufgenommenen Funkspruches zu sein.)

Bei der Konzeptentwicklung zur weiteren Spezifikation der drei zu erstellenden Demonstratoren (Bord, Land, Simulator) arbeitete die DGzRS gemeinsam mit dem Unterauftragnehmer in enger Abstimmung mit den Verbundpartnern zusammen. So waren für die Auswahl der zu verwendenden Hardware-Komponenten neben den in TAP 1.2 identifizierten Anforderungen insbesondere die jeweiligen Einsatzbedingungen zu berücksichtigen: Ein Funkpeilsystem, das an Bord eines Schiffes eingesetzt werden soll (*Borddemonstrator*), muss besondere Anforderungen erfüllen. Zum einen muss es auf der Mastspitze montiert werden können und zum anderen muss das System den maritimen Umweltbedingungen im Hinblick auf mechanische Stabilität und Wasserdichtigkeit gewachsen sein. Zu beachten ist weiterhin, dass die Peilantennen nicht auf der gleichen Höhe wie die AIS- und GPS- Antennen montiert sind, um Reflexionen und Störstrahlungen zu vermeiden.

Für den *Landdemonstrator* sollten zwei Peilanlagen an zwei verschiedenen Küstenstandorten aufgebaut werden. Dabei liefert jeder Peilstandort seine Peilung und AIS-Daten an Bremen Rescue Radio, wo diese von der ARTUS-Anwendung verarbeitet werden. Die Installation von zwei Funkpeilern bietet den großen Vorteil, dass durch die auf diese Weise ermöglichte Kreuzpeilung die Positionsbestimmung eines Funkspruchabsenders wesentlich genauer erfolgen kann. Aus technischer Sicht ist bei der Auswahl geeigneter Standorte⁵ zu beachten, dass eine ausreichende Höhe des Antennenmastes realisiert werden kann. Um möglichst genaue Peilergebnisse zu bekommen, muss die Peilantenne ganz oben platziert werden und darf nicht von anderen metallischen Objekten (z.B. anderen Antennen) verdeckt werden. Außerdem muss ein Geräteschrank von ausreichender Größe und mit entsprechenden Schnittstellen unterhalb des Antennenmastes installiert werden können. Die DGzRS stellte hierzu Informationen zu möglichen Standorten, insbesondere bezüglich eigener Standorte, wie den Funkmast

⁵ Zur Auswahl geeigneter Standorte vgl. genauer die Ausführungen zu AP 4 auf S. 24.

Cuxhaven-Sahlenburg, zur Verfügung. Neben der Verfügbarkeit der erforderlichen Schnittstellen und räumlichen Möglichkeiten zur Installation der notwendigen Hardware waren hier insbesondere auch Fragen der während der Installationsarbeiten notwendigen Abschaltung des vorhandenen Mobilfunkmastes zu berücksichtigen.

Bei der Spezifikation der Demonstrator-Komponenten für den *Schifführungssimulator* war zu beachten, dass hier die Integration des ARTUS-Systems in das geschlossene System der Simulationsumgebung erfolgen musste, das nicht selbständig erweitert werden kann. Die DGzRS sorgte hier für die Kommunikation mit dem Simulatorhersteller, um notwendige Unterlagen bereitstellen zu können.

TAP 1.4 Konzeption & Design der Mensch-Maschine-Schnittstelle

Damit die zu programmierende Software eine praktische Unterstützung für die Einsatzkräfte in ihrem jeweiligen Arbeitskontext darstellen kann, ist es wichtig, dass die Informationen, die durch die Spracherkennungs- und Lokalisierungsalgorithmen bereitgestellt werden, auf der grafischen Benutzeroberfläche (Mensch-Maschine-Schnittstelle) des ARTUS-Systems anwenderfreundlich präsentiert werden. Daher wurde in diesem TAP ein besonderes Augenmerk darauf gerichtet, vor der Übergabe der Systemspezifikation aus TAP 1.3 an den Industriepartner diesbezügliche Erkenntnisse aus dem Endnutzer-Workshop noch einmal gesondert zu extrahieren und zusammenzufassen. Wichtige Aspekte waren neben dem Entwurf der Kartenansicht und der Anzeige der transkribierten Funksprüche auch die Auswahl einzublendender schiffsbezogener, lokalisierungsspezifischer und kommunikationsrelevanter Zusatzinformationen sowie ein Einstellungsmenu für die Peiler.

2.1.2. AP 2 Erfassungssystem

Ziel des zweiten Arbeitspaketes war die Generierung der grundlegenden Daten für die Programmierung des Spracherkennungssystems. Gemeinsam mit dem Unterauftragnehmer und den anderen Verbundpartnern waren drei verschiedene Datensets zu erstellen, die im nachfolgenden Arbeitspaket 3 als Trainingsdaten zu nutzen waren.

Hier waren die DGzRS bzw. der Unterauftragnehmer in alle TAP eingebunden.

TAP 2.1 Systemintegration

Die Arbeitsergebnisse des AP 1 sollten hier in ein PC-basiertes Erfassungssystem überführt werden, welches anschließend zur Generierung eines maritimen Stammdatensatzes (TAP 2.2) und zur Erfassung von Live-Kommunikation bei Bremen Rescue Radio (TAP 3.2) verwendet werden sollte. Zunächst war vom Unterauftragnehmer ein Erfassungssystem zu programmieren, welches dazu dienen sollte, die bei Bremen Rescue Radio eingehenden Funkmeldungen automatisch zu erfassen und zu transkribieren. Diese Software sollte bei BRR installiert werden und den Radio Operatoren die Möglichkeit geben, die maschinell erstellten Transkripte der Funkmeldungen zu überprüfen und manuell zu korrigieren. Die Realisierung des Erfassungssystems verzögerte sich zunächst. Auch bedingt durch Schwierigkeiten bei der Lösung der Schnittstellenproblematiken konnte abschließend nur eine modifizierte und gegenüber der ursprünglichen Planung stark vereinfachte Lösung in der äußeren Form einer „Website“ bereitgestellt werden. Diese erforderte jedoch das manuelle Hochladen jeder einzelnen Audio-datei, was im laufenden Betrieb einen hohen Zusatzaufwand bedeutet. Zuerst war auszuwählen, ob es sich um eine englisch- oder deutsch-sprachige Mitteilung handelt. Während die Sprachmeldung im Hintergrund erkannt wird, erscheint das Transkriptionsergebnis im Textfeld („analyze“). Bei Bedarf kann die verschriftliche Meldung noch verbessert und dann abschließend abgespeichert werden. („save“). Mittels Replay-Funktion kann die Meldung auch mehrmalig wieder angehört werden.

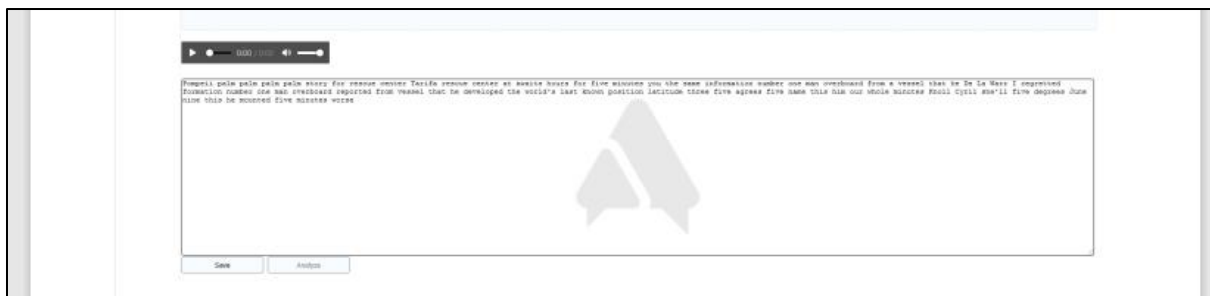


Abbildung 4 Benutzeroberfläche des Erfassungssystems (Entwurf)

Gegenstand der sich anschließenden Teilarbeitspakete 2.2 und 2.3 war die Generierung verschiedener kontextspezifischer Datensets für die Programmierung und das Training des Spracherkennungssystems.

TAP 2.2 Erzeugung von Trainingsdaten (SMCP)

In diesem TAP waren SMCP-Trainingsdaten für das Spracherkennungssystem zu erstellen. Um eine feldspezifische Programmierung zu erreichen, sollte das tiefe neuronale Netz mit Sprachdaten des SMCP-Vokabulars gespeist werden. Außerdem waren beim Unterauftragnehmer bereits vorhandene Datensätze nachträglich zu verrauschen um auf diese Weise Trainingsdaten der gleichen akustischen Güte wie Funksprüche zu erhalten.

Ein neuronales Netz kann ganz allgemein definiert werden als eine „Abbildungsvorschrift ..., [die] eine Menge von Eingaben, deren spezifische Eigenschaften durch sogenannte Eingabevektoren kodiert werden, in eine Menge von Ausgaben, die ebenfalls durch Vektoren beschrieben werden“ (Scherer, 1997, S. 4) wiedergibt. In seiner Schichtstruktur ähnelt es dem menschlichen Gehirn. Die „Tiefe“ eines neuronalen Netzes bezieht sich auf dessen „Architektur“ und gibt Auskunft über die Anzahl der Ebenen bzw. Schichten in denen die eingegeben Daten analysiert werden. Zwischen der Eingabe- und der Ausgabeschicht befinden sich beliebig viele weitere „verborgene“ Schichten. Je tiefer das Netz, also je höher die Anzahl der Schichten, in denen Rechenoperationen stattfinden, desto größer seine Leistungsfähigkeit (Schulz & Behnke, 2012). Tiefe neuronale Netze eignen sich besonders auch für die Erkennung von Mustern zur Identifizierung und Klassifizierung von Sprachsignalen.

Ein „klassisches“ Spracherkennungssystem (vgl. Abb. 5) setzt sich aus drei Komponenten zusammen: Das akustische Modell, die Aussprachedatenbank sowie das Sprachmodell. Im akustischen Modell (*acoustic model*) werden Audiomerkmale einzelnen Lauten (Phonemen) zugeordnet. Dabei geht es darum, die „wahrscheinlichsten“ Phoneme aus akustischen Merkmalen zu ermitteln, die aus dem Audiomaterial extrahiert wurden. Die Aussprachedatenbank (*pronunciation model*) ist nichts anderes als ein „Wörterbuch“, das die Laute in Wörter überführt, also Phonemfolgen auf Wörter abbildet. Im Sprachmodell (*language model*) werden dann auf der Basis von statistischen Modellen aus den erkannten Wörtern Wahrscheinlichkeiten von Wortkombinationen ermittelt und auf diese Weise kontextbezogen sinnvolle Sätze identifiziert. Die Ergebnisse aller drei Komponenten werden nach einer globalen Suche nach den „wahrscheinlichsten“ geäußerten Sätzen als Transkript mit der Audiodatei kombiniert. Hierbei kommen jeweils verschiedene statistische Modelle zum Einsatz (Jaitly, 2017).

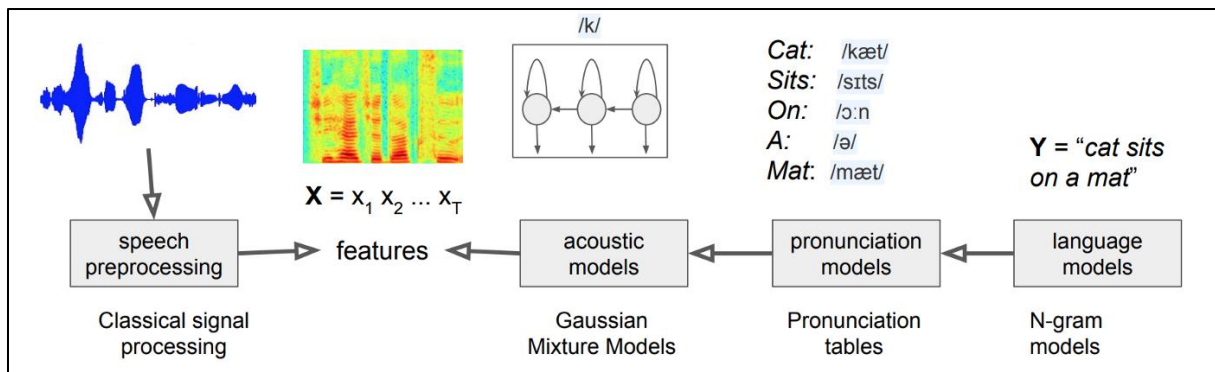


Abbildung 5 Klassisches Spracherkennungssystem⁶

Im Kontext der maritimen Seefunkkommunikation kann die Bedeutung der einzelnen Komponenten wie folgt beschrieben werden: Im *akustischen Modell* kommt vor allem die akustische Güte der verwendeten Trainingsdaten zum Tragen: UKW-Seefunkkommunikation zeichnet sich je nach Übertragungsweg und Störeinflüssen durch einen mehr oder weniger ausgeprägten Verrauschungsgrad aus. Der Spracherkenner muss also in der Lage sein, auch Meldungen zu transkribieren, die infolge von Bandlimitierung, Übersteuerung und sonstigen Störfaktoren nur eine eingeschränkte akustische Güte aufweisen. Für den Aufbau des *Sprachmodells* kommt es auf das verwendete Vokabular sowie die Art und Weise der Kombination von Wörtern zu Sätzen an. Der für das ARTUS-System bereitzustellende Spracherkenner muss also auf die Besonderheiten des im Kontext maritimer SAR-Einsätze verwendeten Begriffe und Wendungen trainiert sein. Hierbei spielen sowohl SMCP-Standardredewendungen als auch die Fülle der in der maritimen Funkkommunikation verwendeten Schiffsnamen und Ortsbezeichnungen eine zentrale Rolle. Schließlich ist zu beachten, dass eine unübersehbar große Zahl an Kommunikationsteilnehmern mit unterschiedlichsten Dialekten und Akzenten auf den Frequenzen des maritimen Seefunks aktiv ist. Dabei sind in den hier relevanten Einsatzgebieten von Nord- und Ostsee nicht nur die in verschiedensten Aussprachevarianten aufgenommenen englischen Sprachmeldungen, sondern auch die nicht selten verwendeten nieder- bzw. plattdeutschen Wendungen zu berücksichtigen. Hier kommt die Bedeutung der *Aussprache-datenbank* zum Tragen.

Neuronale Netze können mit „Experten“ verglichen werden, die sich dadurch auszeichnen, dass sie in der Lage sind, Lösungsstrategien, die sie aus in der Vergangenheit erfolgreich gelösten Problemen erlernt haben, auf neue Probleme anzuwenden. Neuronale Netze eignen sich diese Fähigkeit über das Training mit spezifischen Daten an. Von besonderer Bedeutung für die Leistungsfähigkeit des neuronalen Netzes ist - neben der Menge⁷ - vor allem auch die Qualität der Trainingsdaten (Scherer, 1997, S. 5 f.). Für die Entwicklung eines neuronalen Netzes zur Spracherfassung und Transkription von UKW-Seefunkkommunikation bedarf es somit der Bereitstellung bzw. Erzeugung geeigneter Trainingsdaten, um eine hohe Performanz in der Anwendungsphase sicherzustellen.

Zur Erzeugung von Trainingsdaten für das Sprachmodell hat der Unterauftragnehmer einen Datensatz (ca. 2 Std.) aus dem CML-Simulator um weiteres SMCP-Vokabular ergänzt. Bei diesem Datensatz handelt es sich um unter Laborbedingungen aufgenommene Kommunikation auf der Basis von SMCP typischen Wörtern, Wendungen und Sätzen, die anschließend verschriftlicht wurden. Aufgrund der Laborbedingungen weist der Simulator-Datensatz gute akustische Eigenschaften auf und ist insofern geeignet als Grundlage für das Sprachmodell. Außerdem wurden ca. 2000 typische Phrasen und Formulierungen aus den SMCP extrahiert

⁶ Quelle: (Jaitly, 2017), slide 6.

⁷ Dieser Aspekt war in dieser Projektphase noch von großer Bedeutung, sollte sich gegen Projektende jedoch relativieren. Vgl. S. 20.

und das Sprachmodell des vorhandenen IAIS-Spracherkenners um diese Phrasen und Begriffe ergänzt.

Als weiteres Set wurde ein beim Unterauftragnehmer vorhandener Datensatz digital verrauscht. Hierfür wurden „ideale“ Daten, also Daten von hoher akustischer Qualität, künstlich übersteuert und verrauscht (additive Störgeräuschüberlagerung), so dass sie in ihren akustischen Eigenschaften den auf UKW empfangenen Funksprüchen entsprechen. Die Verrauschung von Daten stellt eine Vergrößerung des Datensatzes („Data Augmentation“) durch Hinzufügen von leicht modifizierten Kopien bereits bestehender Daten dar. Auf diese Weise konnte der Einfluss der für den UKW-Seefunk typischen Störgrößen im akustischen Modell des Spracherkenners besser kontrolliert werden. In dieser Projektphase fußte das Training des Spracherkenners auf der von *Daniel Povey* et al. 2009 im Rahmen eines Workshops an der Johns Hopkins University und 2011 erstmals beschriebenen open-source State of the Art *Kaldi*-Spracherkennungstechnologie (Povey, Ghoshal, & Boulianne, 2011).⁸

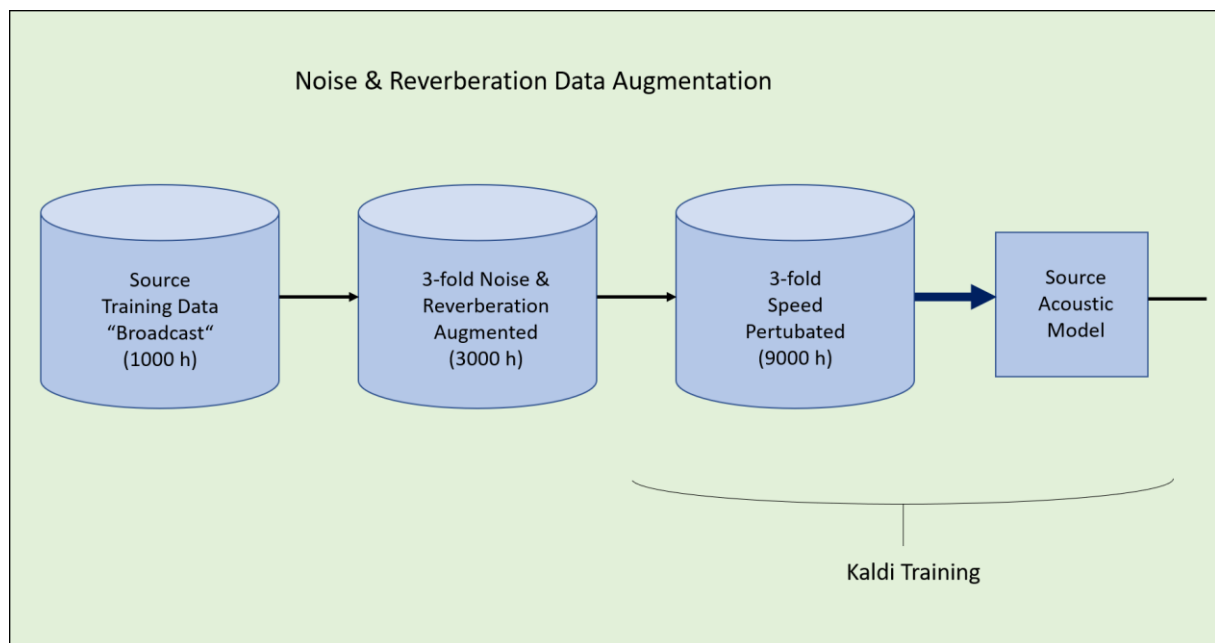


Abbildung 6 Data Augmentation durch Verrauschung⁹

Die digitale Verrauschung lieferte gute Ergebnisse, so dass die als Rückfallebene beim Industriepartner vorgesehene analoge Verrauschung über eine Funkstrecke mit zwei Mikrofonen und damit „reale“ Simulation des Rauschens nicht erforderlich wurde. Im Gegensatz zu einer nur in Echtzeit durchführbaren analogen Verrauschung der Daten konnte durch die erfolgreiche digitale Verrauschung eine deutliche Zeitersparnis erzielt werden.

Der Unterauftragnehmer unterstützte das Forschungshaben zusätzlich durch die Bereitstellung und die Wartung einer auf IAIS-internen Servern rechnenden Demo-Instanz des Spracherkenners.

TAP 2.3 Erzeugung von Trainingsdaten (Bremen Rescue Radio)

In diesem TAP waren Audiodaten aus dem bei BRR eingehenden Funkverkehr des UKW-Kanals 16 manuell zu verschriftlichen.

In TAP 2.3 wurden auf der Basis von realen Funksprüchen von Kanal 16 (UKW-Seefunk), die von BRR zur Verfügung gestellt wurden, Trainingsdaten generiert. Hierbei handelt es sich um

⁸ Zu den Gründen und der Vorgehensweise des im weiteren Projektverlauf vollzogenen Technologiewechsels vgl. die Ausführungen zu AP 3.2 auf S. 19 ff.

⁹ Quelle: IAIS.

Audiodaten, die einerseits SMCP-typisches Vokabular enthalten und somit der Erweiterung des lexikalischen Datensets (Sprachmodell) dienen. Da es sich um Audioaufzeichnungen handelt, die reale akustische Eigenschaften aufweisen, dient dieses Datenset darüber hinaus der Adaption des akustischen Modells des Spracherkenners. Die verwendeten Audiodaten sind zu einem großen Teil Multikanalaufnahmen von BRR. Diese stellen den gleichen Funkspruch von mehreren Relaisstationen aufgezeichnet in unterschiedlichen Aufnahmequalitäten zur Verfügung und können daher noch einen zusätzlichen Beitrag zur Optimierung des akustischen Modells liefern. So wurden die wichtigsten akustischen Eigenschaften dieser UKW-Daten analysiert und benutzt, um einen bestehenden, umfangreichen Trainingsdatensatz anzureichern. Auf diese Weise kann das akustische Modell im nächsten Schritt noch besser an die speziellen Gegebenheiten von UKW-Funk angepasst werden, was die Spracherkennung auch unter realistischen, d.h. nicht idealen Bedingungen verbessern soll.

Die zu transkribierenden Funkmeldungen wurden durch manuellen Export von Kommunikationsdaten aus dem Audioaufzeichnungssystem von BRR gewonnen. Nach der Sammlung der Audiodateien, die zunächst nach englischen und deutschen Meldungen separiert wurden, sind gemeinsam mit dem CML und dem IAIS Transkriptionsregeln aufgestellt und in einem Regelheft zusammengefasst und fortlaufend ergänzt und verfeinert worden.¹⁰ Auf der Grundlage der aufgestellten Regelungen erfolgte die manuelle Verschriftlichung der Audiodateien(.wav) in entsprechenden Textdateien (.txt). Vor der Verschriftlichung erfolgte eine Sichtung des Materials hinsichtlich der Erfüllung bestimmter Mindestanforderungen. So fand ein Audiotranskript nur dann Eingang in das Trainingsmaterial, wenn mindestens 70% der Meldung transkribiert werden konnte. Ein spezielles Augenmerk war auch auf den Umgang mit Pausen zu richten. Um sicherzustellen, dass die Meldung vom Spracherkenner als eine eigenständige Nachricht erkannt wird, durften die Pausen zwischen einzelnen gesprochenen Worten nicht länger als eine vorher festgelegte Zahl von Sekunden betragen. Andernfalls waren die Audiodateien entsprechend zu bearbeiten. Je nachdem, ob es sich um eine englisch- oder eine deutschsprachige Sprachmeldung handelte, wurde der Transkription die Kennung „EN“ (für Englisch) oder „DE“ (für Deutsch) vorangestellt.

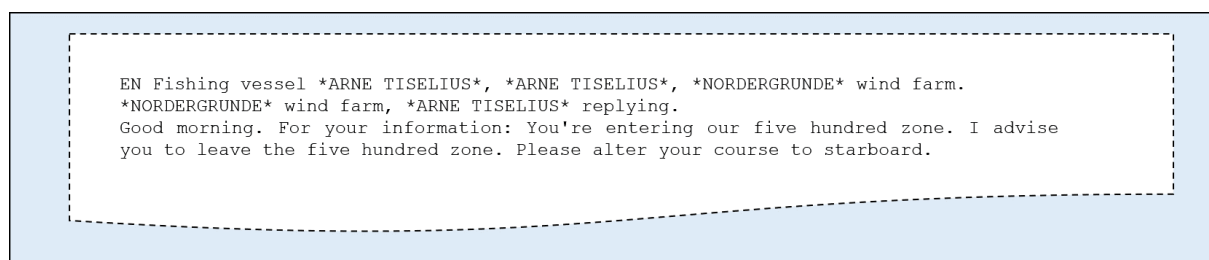
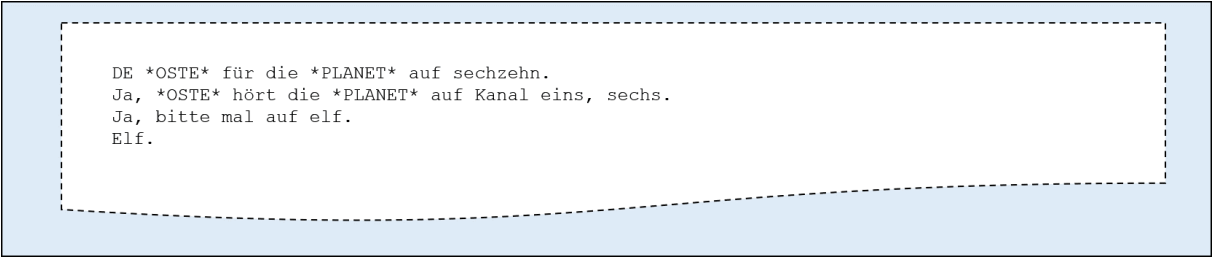


Abbildung 7 Beispiel für ein Transkript (Routinekommunikation) in englischer Sprache

Ein wichtiger Aspekt war auch der Umgang mit lexikalisierten *Anglizismen*, wie z.B. „Standby“, oder „Update“. Wurde in einer ansonsten rein deutschsprachigen Meldung ein solcher Anglizismus verwendet, war die entsprechende Datei dennoch als „clean“ DE zu kodieren.

¹⁰ Diese Transkriptionsregeln legen u.a. fest, wie das Transkript aufgebaut ist, welche Besonderheiten hinsichtlich der Schreibweise z.B. von (Schiffs-) Namen, Zahlen etc. zu berücksichtigen sind, wie unverständliche Worte bzw. Passagen oder Wortabbrüche, Versprecher, Füllworte etc. zu kodieren sind etc.



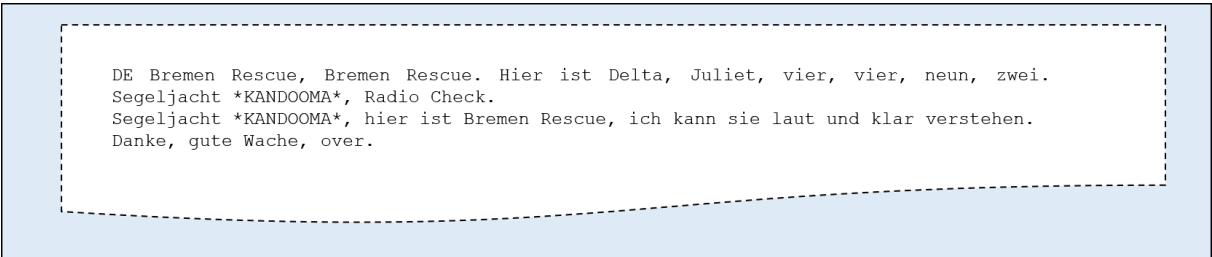
```

DE *OSTE* für die *PLANET* auf sechzehn.
Ja, *OSTE* hört die *PLANET* auf Kanal eins, sechs.
Ja, bitte mal auf elf.
Elf.

```

Abbildung 8 Beispiel für ein Transkript (Routinekommunikation) in deutscher Sprache

Zusätzlich wurden alle in den manuell transkribierten Meldungen verwendeten Anglizismen ergänzend zum Regelheft erfasst und dokumentiert und dem IAIS zur Verfügung gestellt. Im Falle einer gemischtsprachigen Datei, richtete sich die Kategorisierung in „EN“ oder „DE“ nach der in der Sprachmeldung überwiegend verwendeten Sprache.¹¹ Diese „Mixed“ Meldungen wurden in separaten Ordnern abgelegt.



```

DE Bremen Rescue, Bremen Rescue. Hier ist Delta, Juliet, vier, vier, neun, zwei.
Segeljacht *KANDOOMA*, Radio Check.
Segeljacht *KANDOOMA*, hier ist Bremen Rescue, ich kann sie laut und klar verstehen.
Danke, gute Wache, over.

```

Abbildung 9 Beispiel für ein gemischtsprachiges Transkript (DE-mixed)

Wie noch näher auszuführen sein wird,¹² stellte sich die Differenzierung in „EN“ und „DE“ und somit auch die Separierung von „Mixed“ Meldungen gegen Ende des Projektes als nicht mehr notwendig heraus.

Unabhängig von der Kategorisierung erfolgte die Zuordnung von Text- zu Audiodatei durch die Verwendung eines identischen Dateinamens (Zeitstempel und Nummer der Relaisstation auf der die Meldung aufgezeichnet wurde). Ein Datensatz besteht also immer aus einer Audio- und einer Textdatei mit jeweils identischen Dateinamen.

Während sich die Datensammlung in der ersten Hälfte der Projektlaufzeit vorwiegend auf englische Daten konzentrierte, wurden erst im weiteren Verlauf auch deutschsprachige Daten transkribiert. Der noch vergleichsweise geringe Datenbestand mit Funkkommunikation in deutscher Sprache konnte infolge der pandemiebedingten Absagen aller großen Seenotrettungsübungen und Simulator-Exercises (bei denen die Kommunikation aufgezeichnet werden sollte) nicht wie geplant aufgefüllt werden. Als Alternativlösung wurde beschlossen, die Kommunikation realer Seenotrettungseinsätze, die vornehmlich in deutscher Sprache aber z.T. gemischt englisch-deutsch koordiniert wurden, für die weitere Nutzung im ARTUS-Projekt zu sichern. Im Zuge der manuellen Transkription des Audiomaterials aus realen Rettungseinsätzen wurde deutlich, dass zusätzliche Regelungen bezüglich der Abgrenzung von DE- und EN-Kommunikation hinzugefügt und die Kategorisierung in DE- und EN-Daten fortlaufend überprüft werden mussten. Außerdem erfolgte eine intensive Auseinandersetzung bezüglich des Umgangs mit platt- bzw. niederdeutschen Besonderheiten sowie umgangssprachlichen Wendungen in den Transkripten. Zur Sicherstellung der Einheitlichkeit bei der Transkription wurden entsprechende Regelungen im Detail dokumentiert und dem IAIS zur Verfügung gestellt.

¹¹ Dabei wurde z.B. eine DE-Datei auch dann als „Mixed“ kategorisiert, wenn sie nur ein einziges englisches Wort enthielt, das nicht zu den lexikalisierten Anglizismen zu zählen war. Der umgekehrte Fall (ein oder mehrere deutsche Worte in einer EN-Datei) konnte aufgrund fehlender praktischer Relevanz vernachlässigt werden.

¹² Vgl. unten S. 21.

Das TAP 2.3 wurde über den ursprünglich geplanten Zeitraum fortgeführt, so dass der Bestand an transkribierten Kanal-16-Funkmeldungen aus realen Rettungseinsätzen sowie aus Routinekommunikation auf ca. 112 Stunden englisch- und 20 Stunden deutschsprachiges Audiomaterial ausgeweitet werden konnte. Für das Training des Spracherkenners sollte sich die Unterscheidung in deutsch- und englischsprachige Funkmeldungen im weiteren Verlauf als entbehrlich erweisen. Wie oben bereits angedeutet und im Rahmen der Darstellung der Ergebnisse von TAP 3.2 noch zu zeigen sein wird,¹³ wurde durch einen Technologiewechsel u.a. auch die Transkription von englischer und deutscher Funkkommunikation ohne Notwendigkeit einer vorherigen Kategorisierung möglich. Mit anderen Worten: Der ARTUS-Spracherkennner ist in der Lage, aufgenommene Funkkommunikation unabhängig davon, ob es sich um englische, deutsche oder auch gemischtsprachige Meldungen handelt, fehlerfrei zu transkribieren.¹⁴ Die Kategorisierung der Transkripte in deutsche und englische Kommunikation wurde dennoch bis zum Abschluss der Transkriptionsarbeiten beibehalten um einerseits die Konsistenz der Daten zu wahren und andererseits eine mögliche weitere Verwertung der Datenbasis für andere Forschungszwecke zu vereinfachen.

2.1.3. AP 3 Spracherkennung und Programmierung

Im dritten Arbeitspaket erfolgte auf der Grundlage der in Arbeitspaket 2 erstellten Datensätze parallel zum Design und der Implementierung der Senderlokalisierung durch den Industriepartner das Training des Spracherkennungssystems und die Programmierung des tiefen neuronalen Netzes durch den Unterauftragnehmer. Die DGzRS und der Unterauftragnehmer waren in die TAP 3.2 – 3.4. eingebunden

TAP 3.2 Training/Programmierung neuronales Netz

Ziel dieses TAP war es, mit Hilfe der in TAP 2.2 und 2.3 erstellten Trainingsdaten sowie Dokumenten, die SMCP beschreiben, mehrere Spracherkennungssysteme zu trainieren, die Deutsch, Englisch sowie SMCP-Phrasen erkennen.

Den weiteren Ausführungen sei an dieser Stelle bereits vorausgeschickt, dass - wie bereits oben angesprochen - sich im weiteren Projektverlauf abzeichnete, dass auf das Training bzw. die Programmierung *verschiedener* Spracherkennungssysteme (Englisch, Deutsch, SMCP) verzichtet werden konnte und das ARTUS-System mit der Durchführung von multilinguaalem Sprachmodelltraining (Englisch & Deutsch gemischt) in der Lage ist, sprachunabhängig (Englisch/Deutsch) Funkmeldungen zu transkribieren.¹⁵

Das Training des neuronalen Netzes verfolgte zum einen das Ziel auf der Basis der zuvor generierten Daten die Spracherkennung durch die Erweiterung des Sprachmodells um spezielles Vokabular zu verbessern. Zum anderen wurde die Optimierung des Spracherkenners durch die Adaption des akustischen Modells an die akustischen Randbedingungen vorangetrieben.

Für das Training des neuronalen Netzes war es ursprünglich vorgesehen, das in TAP 2.1 erstellte und vorbereitete Erfassungssystem zur weiteren Programmierung für mindestens 12 Monate bei Bremen Rescue Radio zu installieren. Auf diese Weise sollten eingehende Funkmeldungen transkribiert werden und die automatischen Transkriptionen im Erfassungssystem durch die diensthabenden Funker geprüft und im Bedarfsfall korrigiert werden. Über die Erfassung zufälliger Live-Kommunikation und die über die Fehlerkorrektur realisierten Lernschleifen

¹³ Vgl. S. 21.

¹⁴ Im Rahmen einer Kurzdemonstration am Rande der *Mass Rescue Conference* in Göteborg konnten erste Erkenntnisse gewonnen werden, dass der Spracherkennner im aktuellen Reifegrad auch in der Lage ist, weitere Sprachen (hier: Schwedisch) zumindest phonetisch zu transkribieren.

¹⁵ Vgl. hierzu S. 21.

sollte das neuronale Netz trainiert und die Wortfehlerrate (Word Error Rate - WER), welche die Fehlerquote der automatischen Transkription beschreibt und sich als Gesamtzahl von Fehlern, dividiert durch die Gesamtzahl der Wörter in der Referenztranskription berechnet, schrittweise reduziert werden. Als Fehler gelten dabei notwendige Einfügungen, Löschungen und Ersetzungen, die notwendig sind, um von dem Ergebnis der Spracherkennung zum Referenztranskript zu gelangen (Hunt, 1990). Darüber hinaus sollte die Installation des Erfassungssystems einen ersten Feldversuch für die Bedienung der Spracherkenner-Komponente des Landdemonstrators darstellen.

Das bereitgestellte Erfassungssystem konnte nicht, wie vorgesehen, bei BRR installiert werden. Zum einen verfügte die Software nicht im vollen Umfang über die Funktionalitäten, die für einen reibungslosen Einsatz im laufenden Betrieb erforderlich gewesen wären. Durch die fehlende direkte Softwareanbindung hätten die Sprachmeldungen einzeln und manuell in das Erfassungssystem hochgeladen werden müssen. Dies hätte einen erheblichen Zusatzaufwand bedeutet, der im laufenden Betrieb bei Bremen Rescue Radio nicht ohne weiteres hätte realisiert werden können. Zum anderen limitierten im Projektjahr 2020 Umstellungen in der Kommunikations-Software bei gleichzeitigem pandemiebedingten Notbetrieb mit strengen Zugangsbeschränkungen, insbesondere in der Seenotleitung, die Möglichkeiten zur Erfassung von Live-Kommunikation und Audio-Transkription durch die Radio Operatoren. Als Alternativlösung hat der Unterauftragnehmer gemeinsam mit dem Industriepartner RHOTHETA die Möglichkeit geschaffen, die Spracherkennung direkt in den Early Demonstrator zu integrieren. Auf diese Weise konnte die Implementierung der Spracherkenner-Software (Back-End) in ein C++/Qt Front-End zur beispielhaften Nutzung innerhalb der ARTUS-Software realisiert werden. Dabei erfolgten Updates des Spracherkenners jeweils durch die laufende Aktualisierung über einen vom Unterauftragnehmer zur Verfügung gestellten Docker-Container.¹⁶

Das Training des Spracherkenners mit UKW-Audiodaten konnte durch die Fortführung des TAP 2.3 über den ursprünglich geplanten Zeitraum hinaus im Wege der laufenden Zuführung zusätzlichen Datenmaterials an den Unterauftragnehmer sichergestellt werden.

Machine-Learning-Modelle wie die hier zur Spracherkennung verwendeten Neuronalen Netzwerke werden auf Korpora annotierter, d.h. transkribierter, Daten trainiert und evaluiert. Bei der Evaluation werden bereits trainierte Modelle auf sogenannten Test-Daten angewandt.

Corpus-Name	Sprache	train/test-set	Anz. Sprachnachrichten	Anz. Stunden	Anz. Wörter	Anz. Buchstaben	Audio-Dauer in sec 25% Perzentil	Audio-Dauer in sec 75% Perzentil
artus_en_trainset	en	train	21719	111.993	11958	75	12.220	20.890
artus_en_testset	en	test	121	0.579	711	62	12.220	23.440
artus_de_trainset	de	train	3258	16.345	7691	72	10.690	21.592
artus_de_testset	de	test	37	0.219	625	0	11.200	25.990

Abbildung 10 Aufteilung der ARTUS-Daten in Trainings- und Testdaten¹⁷

Für die Evaluation der durch das Spracherkennungssystem generierten Transkripte wurde aus dem Bestand der Trainingsdaten jeweils ein kleinerer Teil (1-2 Std.) als Test-Set separiert und nicht für das Training, sondern für die objektive Beurteilung der Transkriptionsqualität mittels WER herangezogen.

¹⁶ (Docker-) Container sind – vereinfacht ausgedrückt – standardisierte Einheiten von Software, die es Entwicklern erlauben, eine (in der Entwicklung befindliche) Anwendung von ihrer Umgebung zu isolieren, um sie in einer anderen Umgebung zu integrieren bzw. zu testen, vgl. <https://docker.com>

¹⁷ Quelle: IAIS

Beispiel:

aligned_diff	SICHERUNGSBOOT MUNSTER FÜR DEN FORSCHUNGSKUTTER LITTORINA_ SICHERUNGSBOOT MONSDAR FÜR DEN FORSCHUNGSKUTTER LI_TORINAU
--------------	--

Abbildung 11 Beispiel für Fehler im Vergleich zum Referenztranskript¹⁸

Ziel ist es, eine aussagekräftige Abschätzung treffen zu können, wie gut trainierte Modelle ihr „gelerntes Wissen“ auf reale Daten anwenden können und auch auf „neue“, in den Trainingsdaten nicht enthaltene Muster generalisieren. Hierzu ist es notwendig, eine möglichst „realitätsnahe“ Aufteilung der annotierten Daten in Trainings- und Testdaten (Train-Set, Test-Set) vorzunehmen. Eine zufällige Zuteilung zu Train-/Test-Set wird den ARTUS-Daten nicht gerecht, da die von unterschiedlichen Relaisstellen in verschiedenen Qualitäten und z.T. mehrmals wiederholten Funksprüche textuell stark redundante Informationen enthalten können.

Um möglichst verschiedene Funksprüche im Train-Test-Set zu haben, wurde zunächst eine Gruppierung nach Ähnlichkeiten vorgenommen und anschließend (nach manueller Inspektion) eine Zuteilung nur von ganzen Gruppen in Train- oder Test-Set durchgeführt. Es ließen sich etwa 20.000 englische Transkripte in etwa 4.000 Gruppen aufteilen. Etwa 1.000 Gruppen bestehen nur aus einem einzelnen Transkript, 56 bestehen aus acht Transkripten und die größte Gruppe umfasst 1.569 Transkripte. Bei Gruppen mit einer hohen Transkriptanzahl handelt es sich u.a. um wiederkehrende Standard-Sicherheitsmeldungen, die in immer ähnlichem Wortlaut mehrmals täglich aber von unterschiedlichen Sprechern und in unterschiedlicher Qualität gesendet werden.

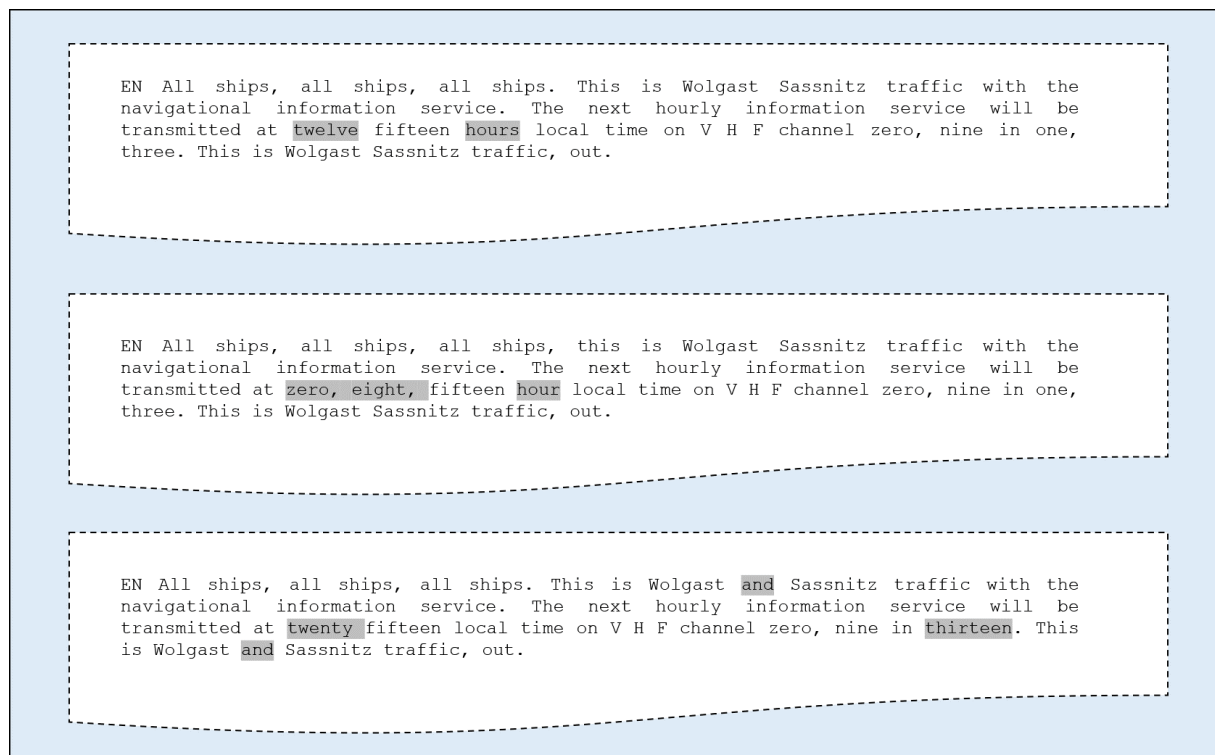


Abbildung 12 Beispiel für quasi-redundante Textfolgen (Standardsicherheitsmeldung)

Gruppen, die jeweils nur eine Meldung enthalten, deuten auf Kommunikation aus Rettungseinsätzen hin, die einsatzspezifische einmalige Textfolgen enthalten und somit keine Redundanzen aufweisen.

¹⁸ Quelle: IAIS

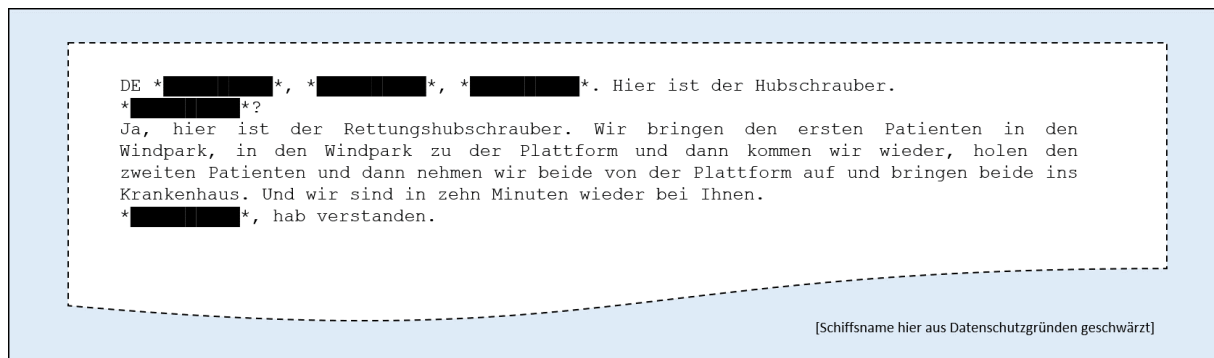


Abbildung 13 Beispiel für einsatzspezifische Textfolge

Die WER des Basismodells des Spracherkenners war vor der Zuführung der domänenspezifischen Trainingsdaten mit über 80% noch sehr hoch. Durch das Training mit den in AP 2 generierten Datensätzen konnte die Qualität des Spracherkennungssystem schrittweise verbessert werden:

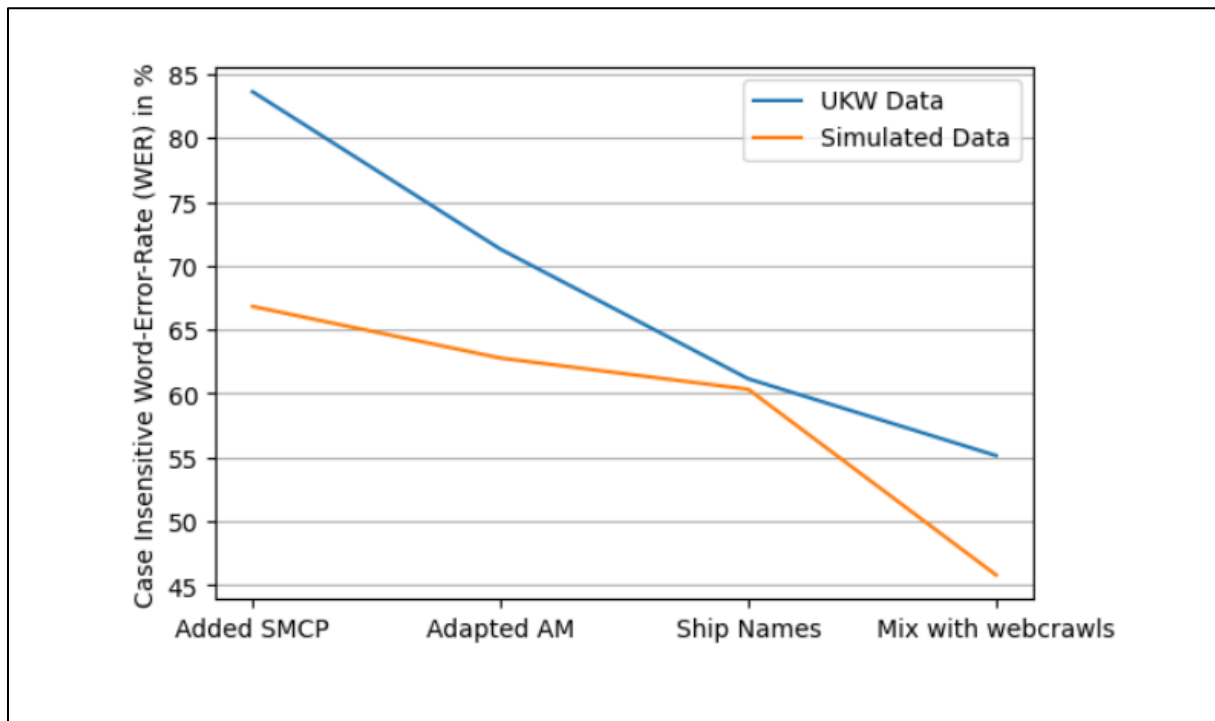


Abbildung 14 Entwicklung der WER in der ersten Projektphase¹⁹

Die WER Entwicklung verlief dabei für die UKW-Daten und die Simulator-Daten ähnlich, unterscheid sich im Detail aber wie folgt:

Ergänzung von SMCP-Vokabular (Added SMCP):

Während sich die WER bei den Simulator-Daten nur leicht verbesserte, da diese bereits gezielt mit SMCP-Vokabular ergänzt wurden, fällt die Verbesserung bei den UKW-Daten etwas deutlicher aus.

Adaption des Akustischen Modells (Adapted AM):

Durch die Adaption des akustischen Modells ergibt sich bei den Simulator-Daten kaum eine Verbesserung, da das Sprachmodell hierdurch nicht verändert wird. Bei den realen UKW-Daten ist eine stärkere Verbesserung der Erkennungsqualität zu beobachten, da durch die

¹⁹ Quelle: IAIS

Adaption den akustischen Rahmenbedingungen der Funkkommunikation (Rauschen, Übersteuerung) Rechnung getragen wird.

Schiffsnamen (Ship Names):

Die korrekte Identifizierung der Schiffsnamen stellt eine besondere Herausforderung an die Spracherkennung dar. Aus diesem Grund wurde das Sprachmodell auf der Grundlage der Auswertung von AIS-Daten (Nord- und Ostsee) um eine Liste mit wahrscheinlichen Schiffsnamen ergänzt. Dies führte auf beide Datensätze bezogen zu einer weiteren Verbesserung der WER.

Weitere Quellen aus dem Internet (mix with webcrawls)

Eine weitere Verbesserung der Erkennungsqualität in beiden Datensätzen konnte schließlich durch die Zuführung von zusätzlichen domänenspezifischen Quellen erzielt werden. Neben Listen mit Häfen, Ortsnamen, Abkürzungen, Fachtermini gehörten dazu auch Textquellen mit realer/gesprochener Sprache. Letzteres ist unter zwei Gesichtspunkten von Bedeutung: (1) Auch wenn das SMCP-Vokabular die Grundlage für die Funkkommunikation auf See darstellen soll, ist in der Alltagspraxis zu beobachten, dass meist *nicht* streng nach den SMCP-Regeln kommuniziert wird. Zwar ist Kanal 16 ausschließlich als Anruf- und Notrufkanal zu nutzen, in der Praxis findet jedoch häufig auch zusätzliche Kommunikation statt, die nicht immer den strengen SMCP-Regeln folgt.

(2) Ein einfaches Hochladen von Listen (Namen, Begriffe) kann das Sprachmodell nur zum Teil verbessern, da der Spracherkenner auf diese Weise nicht die Verwendung der Wörter im Kontext lernt. Hierzu ist es notwendig, dass dem Sprachmodell durch das Training auch Wendungen und ganze Sätze zugeführt werden. Insofern ist an dieser Stelle auch nochmals auf die Bedeutung der Fortführung von TAP 2.3 über den zunächst geplanten Zeitraum hinaus hinzuweisen.

Neben dem Training des englischen Sprachmodells wurde mit der Ausweitung des Bestandes an deutschsprachigen Trainingsdaten auch damit begonnen, den Spracherkenner mit deutschen Aufnahmen zu evaluieren. Bei all diesen Tätigkeiten unterstützte die DGzRS den Unterauftragnehmer neben der laufenden Lieferung von Trainingsmaterial aus TAP 2.3 durch entsprechende Recherchen und Dokumentationen.

Trotz der beschriebenen Maßnahmen im Rahmen der kontinuierlichen Integration des angepassten Lexikons in die Demonstrator-/Produktiv-Software, lag die WER im März 2021 immer noch weit über dem angestrebten Wert von <20%. So entschloss sich der Unterauftragnehmer zu einem grundlegenden Technologiewechsel, weg von der konventionellen „Kaldi“-Technologie hin zur nächsten Generation der automatischen Spracherkennung, zum sogenannten „Wav2Vec-Modell“:

Im Jahr 2017 wurde durch *Vaswani et al.* eine neuartige neuronale Netzwerkarchitektur vorgestellt (Vaswani & Parmar, 2017). Grundidee dieser ist, dass sequenzielle Signale nicht mehr Schritt für Schritt (sequenziell), sondern parallel und somit sehr viel effizienter verarbeitet werden. Zusätzlich zum Effizienzgewinn ermöglicht diese neuronale Netzwerkarchitektur, dass das Modell zeitliche Abhängigkeiten in Zeitreihen besser erfassen und modellieren kann. Diese Netzwerkarchitektur basiert auf dem sog. „Attention-Mechanism“, der erstmals 2015 von (Bahdanau, Cho, & Bengio, 2015, S. 4) vorgestellt wurde und vereinfacht ausgedrückt die komplexitätsreduzierende Eigenschaft beschreibt, sich auf bestimmte Informationen zu fokussieren, um das große Ganze zu erfassen. Damit lehnt sich diese Art der neuronalen Netzwerkarchitektur an Mechanismen der menschlichen Wahrnehmung an, nach denen sich die Aufmerksamkeit unbewusst und intuitiv auf bestimmte Elemente richtet, um das Gesamtbild zu erfassen, z.B. auf einzelne Worte bei der Übersetzung eines Satzes aus einer anderen

Sprache. Ein auf einem solchen Aufmerksamkeitsmechanismus beruhendes Modell ist somit insbesondere auch in der Lage kontextuelle Abhängigkeiten stärker zu berücksichtigen.

Die auf dem Aufmerksamkeitsmechanismus basierende Architektur wurde zunächst als „Transformer“ bezeichnet, da sie zuerst im Bereich der maschinellen Übersetzung (d.h. sowohl Eingabe als auch Ausgabe erfolgen in schriftlicher Form) zum Einsatz kam, und hier eine *Transformation* (Übersetzung) von Quell- zu Ziel-Sprache durchgeführt wird. Im Jahr 2018 wurde diese Architektur mit dem Paradigma des „Selbstüberwachten Lernens“ (self-supervised learning) „verheiratet“. Selbstüberwachtes Lernen bezieht sich auf die Fähigkeit eines Systems selbstständig, d.h. ohne menschliches Zutun, Trainingsmaterial aus vorhandenen Daten zu generieren. Es entstand „BERT“, „Bidirectional Encoder Representations from Transformers“ (Devlin, Chang, Lee, & Toutanova, 2019), ein Modell des Natural Language Processing (NLP) zur kontextuellen Sprachverarbeitung, welches einen neuen State of the Art begründete. NLP ist im Bereich der Computerlinguistik verortet und befasst sich mit der Analyse und Verarbeitung von Text in natürlicher Sprache und soll maschinelle Systeme in die Lage versetzen, menschliche Sprache zu „verstehen“ in der Weise, dass es ihnen möglich ist mit Menschen zu kommunizieren (Otter, Medina, & Kalita, 2019). „Bidirectional“ bedeutet, dass BERT den eingegebenen Text nicht sequenziell also „von links nach rechts“ verarbeitet, sondern in alle Richtungen und damit alle Wörter einer Eingabefolge gleichzeitig verarbeitet. Durch die damit verbundene Erfassung der Wörter in ihrem kontextuellen Zusammenhang unter Nutzung von Selbstaufmerksamkeitsmechanismen konnte die Leistungsfähigkeit neuronaler Netze deutlich gesteigert werden (Devlin, Chang, Lee, & Toutanova, 2019), (Perez, 2019).

Aufgrund des Erfolgs im Bereich des NLP wurden die Ideen des „Attention-based self-supervised Transformers“ auf andere Bereiche übertragen: Ende 2019 veröffentlichte eine Forschergruppe der „Facebook AI Research“ den Code für ihre Wav2Vec-Methodik (Baevski, Fuegen, & Auli, 2019). Im Unterschied zu „traditionellen“ Spracherkennungssystemen, bei denen Audiosignale über die Erkennung von Lauten transkribiert werden, lernt die Wav2Vec-Architektur, das rohe akustische Signal direkt auf Zeichen bzw. Buchstaben zu erkennen.

Wav2Vec steht wörtlich für „wave to vector“. Das bedeutet, dass das Audiosignal (wave-Datei) direkt als Vektor und somit als mathematisches Gebilde abgebildet wird.²⁰ Dies ermöglicht die direkte Umwandlung des hochkomplexen Audiosignals in eine Textdatei über mathematische Operationen. Für die Dekodierung sind keine weiteren Verarbeitungsschritte notwendig. Es gibt keine „klare“ Trennung zwischen akustischem- und Sprachmodell. Im einfachsten Fall ist Wav2Vec beides in einem. Es gibt keine Aussprachedatenbank bzw. kein „Wörterbuch“. „Dekodiert“ wird mittels sogenannter CTC (Connectionist Temporal Classification) -Methodik (Graves, Fernández, Gomez, & Schmidhuber, 2006), (Hannun, 2017). CTC ist ein spezieller Algorithmus zum Training tiefer neuronaler Netzwerke, der das Training des Spracherkenners erheblich vereinfacht (Gref, 2022, S. 30 ff.). Die neue Wav2Vec-Technologie erreichte gegenüber bisher veröffentlichten vergleichbaren Spracherkennungstechnologien eine deutlich verringerte WER bei gleichzeitig drastisch reduziertem Bedarf an transkribierten Trainingsdaten (Schneider, Baevski, Collobert, & Auli, 2019).

2020 konnten mit dem verbesserten Wav2Vec2.0-Algorithmus weitere Erfolge im Bereich der akustischen Sprachverarbeitung erzielt werden. Es konnte gezeigt werden, wie sich Spracherkennungsmittels Rohdaten, d.h. untranskribierten Audiodaten „selbstüberwacht“ trainieren lassen, so dass zur Erstellung eines Spracherkenners im Idealfall nur noch wenige Minuten annotiertes Audiomaterial notwendig sind. Das selbstüberwachte Training mit nicht annotierten Daten führte zu einer weiteren wesentlichen Verbesserung für die multilinguale Spracherkennung (Baevski, Zhou, Mohamed, & Auli, 2020). Da das System mit rohen Daten, d.h. ohne

²⁰ Die Ausgabedatei kann dabei als SVG (Scalable Vector Graphics), Post Script oder CSV (Comma Separated Values) -Format ausgegeben werden (Github).

vorheriges „Klassifizieren der Audiodaten“, trainiert werden kann, bedeutet dies einen erheblichen Effizienzgewinn: Eine vorherige Kategorisierung nach verwendeter Sprache entfällt, was insbesondere für die im ARTUS-Projekt hochgradig relevante Mehrsprachigkeit (Englisch, Deutsch sowie ggf. weitere Sprachen) des Trainingsmaterials von Bedeutung ist.

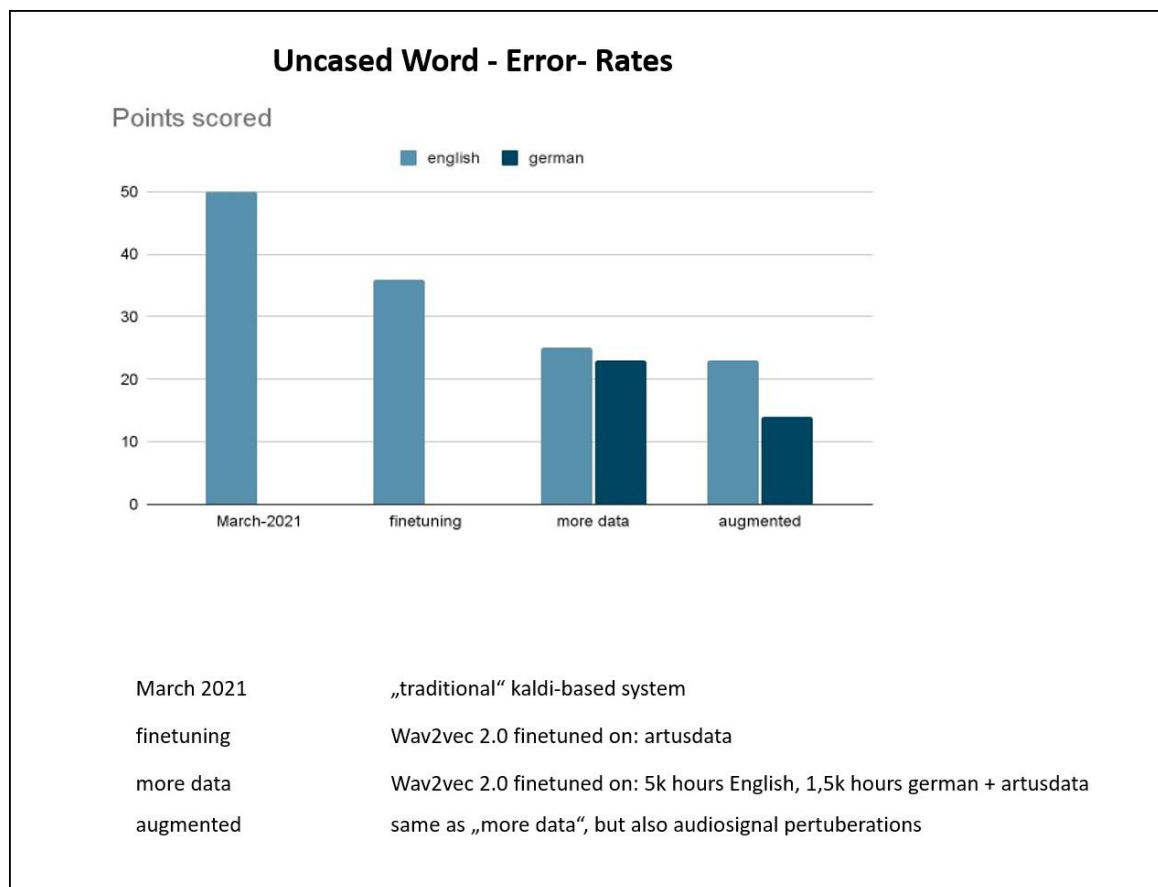


Abbildung 15 Uncased WER nach Technologiewechsel²¹

Die Abbildung zeigt, dass die WER²² vor dem Technologiewechsel noch immer bei 50% lag. Durch das Training des Spracherkenners auf der Basis der *Wav2Vec2*-Technologie durch „finetuning“ auf dem bestehenden ARTUS-Datenset konnte bereits eine Reduzierung der WER um 15 % erreicht werden. Die Vergrößerung des Datensets um zusätzliche Daten und Durchführung eines multilingualen Trainings (Englisch & Deutsch gemischt) führte zu einer weiteren Reduzierung um 10%. Wie bereits angesprochen, ermöglichte der Technologiewechsel nun die Transkription gemischt englisch- und deutschsprachiger Audios.²³ Eine Trennung nach deutsch- und englischsprachigen Daten ist somit nicht erforderlich, die Transkription erfolgt sprachunabhängig. So ist das ARTUS-System in der Lage, problemlos deutsche und englische sowie gemischtsprachige Funksprüche zu erkennen. Damit stellt auch der spontane Wechsel zwischen den Sprachen kein Problem für den Spracherkenner dar. Dies ist von hoher Bedeutung für die Praxistauglichkeit des Systems, da insbesondere bei SAR-Einsätzen in Nord- und Ostsee immer auch in deutscher Sprache kommuniziert wird, insbesondere wenn deutsche Nicht-SAR-Kräfte, z.B. Fischereifahrzeuge oder Sportboote an Einsätzen beteiligt sind.

Insgesamt konnte die Leistungsfähigkeit des Spracherkenners durch die Zuführung weiteren Trainingsmaterials final deutlich verbessert werden. Hier kam der Effekt des

²¹ Quelle: IAIS

²² „Uncased“ WER bedeutet, dass die Berechnung der WER ohne Berücksichtigung von korrekter Groß- und Kleinschreibung erfolgte.

²³ Vgl. S. 21.

Technologiewechsels signifikant zum Tragen, da u.a. durch das Zuführen von nur wenigen Stunden an zusätzlichen Audiotranskriptionen aus einem realen Rettungseinsatz mit entsprechend zusätzlichem einsatzpraktischem Vokabular, die Qualität der Spracherkennung noch einmal auf ein höheres Niveau gebracht werden konnte.²⁴

TAP 3.3 Umsetzung der Schnittstellenkonzepte je Demonstrator

In diesem TAP sollte die Hardware für die drei Demonstratoren zusammengestellt, die Ortungssoftware integriert sowie Korrekturen und Anpassungen vorgenommen werden.

Hierbei hat das IAIS die Verbundpartner bei der Umsetzung der Schnittstellenkonzepte zur Zusammenführung von Spracherkennung und Senderlokalisierung unterstützt. Die Verarbeitung der Audio-Streams, also der Audioausgangssignale des Funkpeilers erfolgt dabei mittels Websocket-Verbindung. Dabei war ein besonderes Augenmerk auf die korrekte Erfassung des Audiosignals und seine Weiterverarbeitung hinsichtlich des Beginns und des Endes der empfangenen Sprachmeldung zu richten. So musste durch geeignete Algorithmen, die von RHOTHETA in enger Zusammenarbeit mit dem IAIS angepasst wurden, sichergestellt werden, dass nicht jedes einzelne gesprochene Wort als eigenständige Nachricht erkannt wird. Hierbei wurden insbesondere die Spannungsschwelle des Eingangssignals sowie die Pausen zwischen den Eingangssignalen berücksichtigt.

Die DGzRS wirkte weiterhin koordinierend bei der Beschaffung der Hardware-Komponenten mit.

TAP 3.4 Implementierung und Test der Mensch-Maschine-Schnittstellen

In diesem TAP war die Mensch-Maschine-Schnittstelle zu implementieren, zu testen und zu optimieren.

In der ursprünglichen Planung sollte das bei BRR zu implementierende Erfassungssystem als Testbett dienen, wobei die DGzRS beabsichtigte, mit Radio Operatoren als potenzieller Nutzergruppe zu unterschiedlichen Zeitpunkten Usability-Tests durchzuführen. Wie bereits mehrfach angesprochen, konnte die Installation des Erfassungssystems in der geplanten Weise nicht stattfinden.²⁵ Stattdessen erfolgte die Integration der Spracherkennungssoftware in den Early Demonstrator. Hierbei arbeiteten das IAIS und der Industriepartner RHOTHETA eng zusammen.

Im Rahmen des Meilensteintreffens am 9. September 2020 konnten die Endanwender einen ersten Eindruck von den Funktionalitäten des ARTUS-Systems gewinnen. Aufgrund der nach wie vor aufrechtzuerhaltenden Maßnahmen des Infektionsschutzes konnte das Meilensteintreffen nur virtuell stattfinden. Der Industriepartner realisierte dazu einen entsprechenden „Versuchsaufbau“ in seinen Räumlichkeiten in Murnau und war damit in der Lage, den Early Demonstrator sowohl mittels emulierter Daten, als auch auf der Grundlage der, von zwei in der Umgebung ausgebrachten AIS-Notsendern, empfangenen realen Daten eindrucksvoll zu präsentieren. Funkmeldungen konnten mittels Mikrofons und Soundkarte eingesprochen werden und in einem separaten Nachrichtenfenster verschriftlicht visualisiert sowie mittels Replay-Funktion erneut abgespielt werden.

Nach einer eingehenden Vorstellung der drei Varianten des Early Demonstrators (Land, Bord, Simulator) und der anschließenden Live-Präsentation des ARTUS-Systems im Labor hatten die potenziellen Endanwender in einer Feedbackrunde Gelegenheit, ihre Eindrücke zu teilen, auf deren Grundlage eine weitere Optimierung und Feinjustierung der MMS erfolgen konnte.

²⁴ Vgl. hierzu auch S. 33.

²⁵ Vgl. S. 16.

2.1.4. AP 4 Erstellung von Demonstratoren

Im Rahmen von AP 4 erfolgte die hard- und softwaretechnische Integration von Spracherkennung und Objektllokalisierung in den Demonstratoren.

Die DGzRS und der Unterauftragnehmer sind im AP 4 vorwiegend koordinierend und unterstützend tätig gewesen. Der DGzRS oblag dabei u.a. das Einholen von Angeboten und die Koordination der Beschaffung der Hardware. Das Einholen der Angebote war am Ende des zweiten Quartals 2021 abgeschlossen. Aufgrund von ausgeprägten Lieferengpässen bei den Hardwarekomponenten konnte das AP jedoch erst verzögert abgeschlossen werden. So stand die für die Erstellung der Bord- und Landdemonstratoren erforderliche Hardware (MOXA-Industrierechner, DELL-High-Performance-Notebooks) erst Anfang 2022 zur Verfügung. Dies führte in der Folge dazu, dass – wie bereits erwähnt – eine zuwendungsneutrale Projektverlängerung in Anspruch genommen werden musste.²⁶

Im Zuge der Softwareintegration waren RHOTHETA und das IAIS in engem Austausch bezüglich verwendeter Datenformate sowie der Auswahl geeigneter Parameter für die einwandfreie Funktionalität der sog. „Voice-Activity-Detection“. Dies betrifft das Erkennen von Start- und Ende einer eingehenden Sprachmeldung und bedurfte einiger Iterationen bis diese kleine, aber sehr wichtige Komponente innerhalb der gesamten Spracherkennungsverarbeitungskette korrekt konfiguriert war.

Eine besondere Herausforderung stellte in der letzten Phase der Systemintegration der Spracherkenner-Komponente in das ARTUS-System die Anpassung des Spracherkenners an die Rechenleistung der verwendeten Hardware dar. Der Spracherkenner kann grundsätzlich auf Rechnern mit unterschiedlicher Leistung zum Laufen gebracht werden, muss sich jedoch an deren jeweilige Rechenkapazität anpassen. Dafür muss die „Chunking-Strategie“, also die Art und Weise der „Bündelung“ bzw. „Stückelung“ der zu verarbeitenden Informationseinheiten (hier: Audiodaten), so angepasst werden, dass der Spracherkenner „echtzeitfähig“ bleibt. So werden die digitalisierten Audiosignale „paketweise“ von der ARTUS-Software an den Spracherkenner geschickt. Sobald die Spracherkenner-Komponente Teile der Sprachmeldung erkennt, wird dieses Ergebnis an das ARTUS-System zurückgeschickt. Das vorläufige Ergebnis wird im Transkriptionsfenster der ARTUS-Oberfläche angezeigt. Im Hintergrund arbeitet der Spracherkenner weiter, so dass sich das komplette Transkriptionsergebnis laufend aktualisiert. Ziel war es, die Geschwindigkeit des Spracherkenners so zu steigern, dass sich das vollständige Transkript innerhalb von wenigen Sekunden aufbaut, die Transkription also in „Echtzeit“ angezeigt wird, um so eine größtmögliche Benutzerfreundlichkeit zu realisieren. Die eigentliche Herausforderung bestand nun darin, die Größe der zu verarbeitenden „Informationspakete“ so zu dimensionieren, dass bei gegebener Hardware/Rechenkapazität Echtzeitfähigkeit erreicht wird. Diese Aufgaben konnte nur durch gemeinsames Experimentieren von IAIS und RHOTHETA realisiert werden. Schlussendlich konnte das System so konfiguriert werden, dass die komplette ARTUS-Software auf einem Rechner laufen kann. Grundsätzlich ist es aber ebenso möglich, Spracherkennung und Senderlokalisierung auf zwei separaten, miteinander verbundenen Rechnern zu betreiben, falls dies die Gegebenheiten im Einsatzkontext z.B. bei Nutzung von vorhandenen Hardwarekomponenten erleichtert oder in bestimmten Kontexten auch nur die Nutzung einer der beiden Komponenten gewünscht sein sollte.

Die Integration des ARTUS-Spracherkenners erfolgte Ende Mai 2022 in der Version 3.0. Als wichtige Anforderung an den Spracherkenner wurde zu Projektbeginn seine Offline-Funktionalität für den Bordbetrieb identifiziert. Im Rahmen einer zukünftigen Weiterentwicklung und Verbesserung der Spracherkenner-Software, wäre es mittels externem Speichermedium (z.B.

²⁶ Vgl. S. 3

USB-Stick) jedoch jederzeit möglich, ein Upgrade zu installieren. Dies ist insbesondere für die Anwendung an Bord notwendig, falls keine stabile Internetverbindung verfügbar sein sollte.

TAP 4.1 Küstenfunkstelle: Schnittstelle Funk & Systemintegration (Soft- und Hardware)

Durch die Realisierung des Landdemonstrators sollten Einsatzmöglichkeiten und Praxistauglichkeit des ARTUS-System in maritimen Rettungskoordinierungszentren oder Küstenfunkstellen, wie BRR oder aber in einem erweiterten/späteren Anwendungskontext auch in den Verkehrszentralen zur maritimen Verkehrssicherung überprüft werden. Durch die Verwendung von zwei Funkpeilanlagen mit zwei simultanen Empfangskanälen wird die Genauigkeit und Zuverlässigkeit der Schiffsortung erhöht. Außerdem ermöglicht dieser Ansatz im Wege der Kreuzpeilung auch dann die Lokalisierung eines Schiffes, wenn dieses kein AIS-Signal ausstrahlt.

Die DGzRS koordinierte die Suche und Auswahl möglicher Standorte für die Landdemonstratoren und war, wie bereits in den Ausführungen zur TAP 1.3 erwähnt, beratend in die Erstellung von Aufbauplänen und Vorbereitung der Anlagen durch den Industriepartner involviert.

Bei der Auswahl geeigneter Standorte eruierte die DGzRS zum einen die Nutzungsmöglichkeiten eigener Anlagen (Funkmast in Cuxhaven-Sahlenburg). Zum anderen führte sie Gespräche mit verantwortlichen Entscheidungsträgern weiterer in Frage kommender Standorte, so z.B. mit der Wasserstraßen- und Schifffahrtsverwaltung (für einen möglichen Standort auf Fehmarn/Ostsee) und dem Helmholtz-Zentrum Geesthacht – jetzt HEREON – (für den möglichen Standort Büsum/Nordsee).

Aus technischer Sicht waren zum einen die Gegebenheiten vor Ort (Antennenhöhe, Möglichkeiten zur Unterbringung der Hardware in einem geeignetem Geräteraum/-schrank) zu berücksichtigen. Hier stellte die DGzRS die notwendigen Informationen zur Verfügung. Zum anderen war der Abstand der beiden in Betracht kommenden Standorte zueinander (Funkreichweite) in die Auswahl einzubeziehen, da das ARTUS-System nur bei Zustandekommen einer Funkverbindung zwischen den beiden Peilern zuverlässig arbeiten kann. Unter Berücksichtigung von Antennenhöhen der sendenden und empfangenden Funkpeiler sowie des Abstands zwischen den jeweiligen Antennenstandorten hat RHOTHETA verschiedene Kombinationsmöglichkeiten berechnet.

Nach Auswertung der Ergebnisse und in enger Abstimmung mit den Projektpartnern fiel die Wahl auf die Standorte Cuxhaven-Sahlenburg und Büsum. Zur Einholung der erforderlichen Aufstell- und Betriebsgenehmigungen führte die DGzRS im Frühjahr 2021 entsprechende Gespräche mit den zuständigen Institutionen.

Durch die zuwendungsneutrale Projektverlängerung konnten die Landdemonstratoren trotz der bereits beschriebenen Lieferengpässe noch im Projektzeitraum vom Industriepartner fertiggestellt werden. Weitere pandemiebedingte Verzögerungen, u.a. bei der Erteilung von Betriebsgenehmigungen an den ausgewählten Standorten sowie wartungsbedingte Restriktionen am Standort Cuxhaven-Sahlenburg verhinderten jedoch den Aufbau der Anlagen innerhalb der Projektlaufzeit. Die Demonstratoren stehen zur Installation bereit und können nach Abschluss der Wartungsarbeiten und Vorliegen aller erforderlichen Betriebsgenehmigungen an den ausgewählten Standorten aufgestellt werden.

TAP 4.2 Mobiler Betrieb an Bord von Einsatzmitteln: Systemintegration (Soft- und Hardware)

Für den mobilen Betrieb an Bord von Einsatzmitteln waren ursprünglich insgesamt drei Demonstratoren geplant. Eine Anlage sollte an Bord eines Rettungskreuzers der DGzRS installiert werden, die beiden anderen Systeme sollten für die DLRG sowie für die Wasserwacht des BRK als assoziierte Partner im ARTUS Verbund bereitgestellt werden.

Im Projektverlauf waren zunächst in enger Abstimmung mit BRK und DLRG geeignete Einsatzmittel für die Installation der Demonstratoren auszuwählen. Es zeigte sich, dass in der Konzeptionsphase einzelne Rahmenbedingungen nicht bekannt waren, die einen sinnvollen Einsatz der Demonstratoren zu einem gewissen Grad limitierten. Hier ist zunächst die erforderliche Mindestgröße des Einsatzmittels zu nennen: Wie bereits im Zusammenhang mit der Systemspezifikation (TAP 1.3) erwähnt, spielt die Montage der Funkpeilantenne in ausreichender Höhe eine zentrale Rolle für die Funktionsfähigkeit des ARTUS-Systems. Dafür ist eine ausreichende Stabilität des aufnehmenden Einsatzmittels (Rettungsboot mit Geräteträger) erforderlich, die sich nur ab einer bestimmten Bootsgröße gewährleisten lässt. Daneben waren die Rahmenbedingungen des potenziellen Einsatzortes zu berücksichtigen: Hier waren zum einen die Nutzung von UKW-Funk bei der Einsatzkommunikation und zum anderen die Verfügbarkeit von AIS-Signalen für die Senderlokalisierung zu nennen.

Bei der BRK-Wasserwacht erfüllt mit Blick auf die baulichen Voraussetzungen ein am Bodensee stationiertes Rettungsboot die Mindestanforderungen am ehesten und befindet sich zudem in relativer räumlicher Nähe zum Industriepartner RHOTHETA. Jedoch stellte sich im Austausch mit den Technikverantwortlichen sowie mit der betreffenden Ortsgruppe des BRK heraus, dass auf dem Bodensee keine AIS-Signale empfangen werden dürfen, weil die entsprechenden zu nutzenden Funkfrequenzen von kritischer Infrastruktur in der Schweiz belegt sind. Außerdem wird auf dem Bodensee vorwiegend über Mobilfunk und Tetra-BOS kommuniziert. UKW-Binnenfunk wird außer vom BRK lediglich von der Wasserschutzpolizei genutzt. Damit könnte das ARTUS-System nur eingeschränkt bzw. mit erheblichem Zusatzaufwand bei der BRK-Wasserwacht eingesetzt werden. Daher entschloss sich der Projektverbund in Abstimmung mit dem assoziierten Projektpartner von einer Integration eines Demonstrators auf dem Einsatzboot des BRK abzusehen. Der Austausch mit dem assoziierten Projektpartner wurde dennoch fortgesetzt, da unabhängig von den Möglichkeiten einer Senderlokalisierung, Interesse an Fragen der automatischen Spracherkennung in der Einsatzkommunikation geäußert wurde.

Auch für den Einsatz beim assoziierten Partner DLRG war zunächst ein geeignetes Einsatzmittel auszuwählen. Baulich kamen hier Rettungsboote sowohl im Raum Berlin-Brandenburg als auch auf der Elbe in Frage. Auf Empfehlung des IuK-Bereiches des DLRG-Bundesverbandes wurde mit Blick auf die technischen Gegebenheiten vor Ort sowie die Art der zu leistenden Einsätze die Integration des ARTUS-Systems auf einem Rettungsboot im Bereich der Elbe am Standort Stade beschlossen. Erste, sehr konstruktive Gespräche mit der Ortsgruppe Stade zur Auslotung der Integrationsmöglichkeit eines ARTUS-Demonstrators an Bord eines Rettungsbootes auf der Elbe ergaben, dass das zur Disposition stehende Einsatzmittel zwar grundsätzlich groß genug ist, jedoch auch verschiedene einsatzpraktische Gegebenheiten zu berücksichtigen waren, die insbesondere bei der Installation der Funkpeilantenne limitierend wirkten: So müsste der hierfür zu nutzende Geräteträger in der Einsatzpraxis häufig umgelegt werden, um unter Brücken durchfahren zu können. Die Montage der Antenne an einer anderen Stelle des Bootes würde den Bereich zur Übernahme von Patienten beschränken und war somit auch keine Option. So entschloss man sich schließlich, den Demonstrator in einem Einsatzleitwagen (ELW) der DLRG zu installieren. Während eines Einsatzes laufen dort alle Informations- und Kommunikationsströme zusammen, so dass vor allem auch hier der Spracherkennung von großem Interesse ist. So wurde ein entsprechendes Konzept für die Integration eines Demonstrators in den ELW entwickelt. Um das System flexibel zu halten und gleichzeitig platzsparend und ohne größere Umbaumaßnahmen arbeiten zu können, entschloss man sich für ein Laptop-basiertes System. Als feste Installationen bleiben die Peilantenne, Kompass und GPS. Während die Genehmigung zum Betrieb eines Seefunkgeräts auf einem Einsatzfahrzeug der DLRG zügig erteilt werden konnte, waren aus technischer Sicht weitere Faktoren bezüglich der Verfügbarkeit von für den Betrieb des ARTUS-Systems notwendigen Daten zu berücksichtigen. So besteht bei der Installation in Landfahrzeugen das Problem, dass ein

stehendes Fahrzeug keine peilrelevanten GPS- Daten empfängt (kein „course over ground“). Darüber hinaus sind mögliche Störeinflüsse durch metallische Objekte bei Verwendung eines magnetischen Kompasses zu beachten. Unabhängig von technischen Fragen führten auch hier pandemiebedingte Restriktionen dazu, dass der Demonstrator nicht innerhalb der Projektlaufzeit realisiert werden konnte. Während der ELW in den einsatzstarken Monaten (Frühjahr bis Herbst) nicht für Umbau- und Installationsmaßnahmen zur Verfügung stand, verhinderten im Winter 2021/22 erneute Kontakt- und Reisebeschränkungen die Aufnahme der notwendigen Arbeiten. Dennoch steht die DGzRS weiterhin mit der DLRG-Ortsgruppe Stade in Kontakt und wird diese, sobald es die Gegebenheiten zulassen, zu einer Borddemonstration des ARTUS-Systems nach Cuxhaven einladen. Von besonderem Interesse für die DLRG bleibt dabei die Spracherkenner-Komponente, jedoch nicht nur beschränkt auf den UKW-Funk, sondern daneben auch erweitert auf den digitalen BOS-Funk.

Die DGzRS ihrerseits entschied sich für die Installation des Borddemonstrators an Bord des Seenotrettungskreuzers ANNELIESE KRAMER. Neben der geografisch günstigen Stationierung in Cuxhaven (räumliche Nähe zum Sendemast Cuxhaven-Sahlenburg sowie zur DLRG in Stade) und der damit verbundenen Reduzierung erforderlicher Reisekosten zwischen den Standorten, bietet die ANNELIESE KRAMER als Seenotrettungskreuzer der 28m-Klasse den Vorteil eines zusätzlichen OSC-Arbeitsplatzes auf der Brücke, so dass sich der ARTUS-Demonstrator hier gut installieren und erproben lässt. In der ursprünglichen Planung war der Borddemonstrator als separater Panel PC konzipiert. Im Zuge der Berücksichtigung des Endanwenderfeedbacks (Stationsleiter und Besatzung), stellte sich jedoch heraus, dass eine solche Lösung aufgrund des nur begrenzt zur Verfügung stehenden Platzangebotes an Bord, besondere Herausforderungen an die Integration hinsichtlich erforderlicher bzw. bereits bestehender Mehrfachbelegungen von vorhandenen Rechnern und Displays darstellen würde. So entschloss man sich auf Wunsch der Besatzung, statt des geplanten Panel-PCs ein Laptop-basiertes System am OSC-Arbeitsplatz zu betreiben. Eine vergleichbare Laptop-Lösung hatte der Industriepartner bereits zu Simulations- bzw. Demonstrationszwecken vorbereitet. Für den Einsatz an Bord wurde dafür die erforderliche Hardware vom Industriepartner konfiguriert und bereitgestellt. Bei der Beschaffung der Hardwarekomponenten führten die bereits erwähnten Lieferengpässe dazu, dass der Borddemonstrator erst im Frühjahr 2022 an die DGzRS übergeben werden konnte. Zur Installation an Bord der ANNELIESE KRAMER koordinierte die DGzRS die Abstimmung zwischen der hauseigenen Funkwerkstatt und dem Industriepartner. Aufgrund der für den Rettungsdienst bis zum Ende der Projektlaufzeit strikt einzuhaltenden Infektionsschutzmaßnahmen konnte auch der Borddemonstrator bis zum 30.06.2022 noch nicht installiert werden und die Aktivitäten zur Hardwareinstallation und -integration vor Ort mussten sich auf vorbereitende/planerische Abstimmungsmaßnahmen beschränken. RHOTHETA hat jedoch auch über die Projektlaufzeit hinaus seine Unterstützung bei der Integration des ARTUS-Systems an Bord des Seenotrettungskreuzers zugesichert, so dass eine Installation zeitnah und sobald es die Rahmenbedingungen zulassen, erfolgen kann.

TAP 4.3 Schiffsführungs-/Kommunikationssimulator: Systemintegration (Soft- und Hardware)

Aufgrund von Schwierigkeiten hinsichtlich der Abstimmung mit dem Hersteller des DGzRS-Simulators bezüglich der Integration und Umrüstung und um den laufenden Schulungs- und Ausbildungsbetrieb nicht zu gefährden, musste vorerst von der Installation eines ARTUS-Demonstrators abgesehen werden.

2.1.5. AP 5 Implementierung und Erprobung im Einsatzkontext

Das fünfte Arbeitspaket diente der Erprobung der Praxistauglichkeit der in Arbeitspaket 4 erstellten Demonstratoren in ihrem jeweiligen Einsatzkontext. Durch ein strukturiertes Feedback der Endnutzer sollten Verbesserungsmöglichkeiten der Demonstratoren ermittelt werden.

Infolge der pandemiebedingten Verzögerungen konnte das AP 5 trotz der zuwendungsneutralen Projektverlängerung nicht wie geplant durchgeführt werden. Aufgrund der noch ausstehenden Installationen der Demonstratoren, war es somit nicht möglich, das ARTUS-System innerhalb der Projektlaufzeit im konkreten Einsatzkontext zu erproben.

Um den Endanwendern dennoch einen Eindruck von der Funktionalität und Leistungsfähigkeit des ARTUS-Systems vermitteln zu können, entschloss man sich im Projektkonsortium zu einer Durchführung von Endnutzer-Workshops, die aufgrund der weiterhin aufrechtzuerhaltenden Infektionsschutzmaßnahmen nur virtuell durchgeführt werden konnten.

Insgesamt wurden im Zeitraum von März bis Mai 2022 mehrere Online-Workshops mit Teilnehmern aus den Bereichen Endanwender Bordbetrieb, Land und Simulator sowie Zulassungsbehörde (Bundesamt für Seeschifffahrt und Hydrographie, BSH) durchgeführt. RHOTHETA realisierte dazu inhouse eine Simulationsumgebung, die es ermöglichte, das ARTUS-System im Rahmen von Online-Meetings via MS Teams zu präsentieren. Die Workshop-Teilnehmer hatten auf diese Weise die Möglichkeit, die Funktionsweise über die online geteilte ARTUS-Benutzeroberfläche akustisch und visuell „live“ am Bildschirm mitzuverfolgen sowie die Spracherkenner-Komponente selbst unmittelbar zu testen. Mittels des eigens von RHOTHETA entwickelten speziellen Versuchsaufbaus war es möglich, den Einsatz des ARTUS-Systems an Bord eines Rettungskreuzers (hier SK ANNELIESE KRAMER) zu simulieren. Kurs und Geschwindigkeit des Seenotkreuzers ließen sich in der Simulationsumgebung steuern. In der Präsentation sollte gezeigt werden, wie die ANNELIESE KRAMER einen Funkspruch empfängt, dessen Sender mittels des ARTUS-Systems lokalisiert und identifiziert wird²⁷, während die eingehende Funkmeldung zeitgleich transkribiert wird.

Zur Simulation der Funkpeildaten drückte der Versuchsleiter eine entsprechend eingerichtete und mit der Soundkarte des ARTUS-Rechners verbundene PTT-Taste, während er zeitgleich einen Funkspruch in das verbundene Mikrofon einsprach.²⁸ Die Anzeige der Schiffe auf einer OpenSeaMap-Karte erfolgte über einen internen AIS-Server²⁹. Die Endnutzer konnten den Vorgang der Einblendung der Peillinien zur Objektlokalisierung, die Berechnung der Erkennungswahrscheinlichkeiten („Detection Likelihood“) sowie den Aufbau der Transkriptanzeige während des Funkspracheingangs unmittelbar auf der ARTUS-Benutzeroberfläche mitverfolgen. Auch die Möglichkeit, verschiedene aufgezeichnete Funksprüche jederzeit erneut abspielen zu können, konnte demonstriert werden. Im Anschluss hatten die Teilnehmer die Möglichkeit, selbst einen beliebigen Funkspruch einzusprechen, der mittels Audioübertragung via MS Teams an das ARTUS-System übermittelt wurde. Um einen Ausblick auf das weitere Anwendungsspektrum zu geben, wurde in einer zweiten Versuchsanordnung durch Simulation eines weiteren mit dem ARTUS-System ausgestatteten Seenotkreuzers die Funktionsweise der

²⁷ Wobei durch das ARTUS System mehrere mögliche („wahrscheinliche“) Absender identifiziert werden können: Der Objektlokalisierungsalgorithmus berechnet jedoch für alle auf oder in der Nähe des Peilstrahls liegenden Objekte (Schiffe) die Erkennungswahrscheinlichkeiten, so dass am Ende das Schiff mit der höchsten Wahrscheinlichkeit als Absender angenommen wird.

²⁸ RHOTHETA hatte bereits im Frühjahr 2021 im Zuge der Installation eines Funkpeilernetzwerks im Auftrag des Wasserstraßen- und Schifffahrtsamts auf der Insel Fehmarn die Möglichkeit, einzelne Funktionalitäten des ARTUS-Systems zu testen. So konnte im Rahmen eines Feldversuches nachgewiesen werden, dass der Empfang des Audio-Funksignals gleichzeitig mit der Peilanzeige (Einblendung der Peillinien) und der Anzeige der Senderlokalisierung (unter Angabe der Erkennungswahrscheinlichkeit) erfolgt.

²⁹ Da für die Simulation bei RHOTHETA in Murnau keine Echtzeitdaten genutzt werden konnten, wurden von der Firma *Jakota Cruise Systems* bereitgestellte historische AIS-Daten aus dem Seegebiet um Cuxhaven in den Versuchsaufbau eingelesen.

Kreuzpeilung (Triangulation) und damit das Potenzial der Vernetzung von Peilern eindrucksvoll demonstriert.³⁰ Durch Kreuzpeilung ist es möglich, Wasserfahrzeuge auch dann mit hoher Wahrscheinlichkeit zu lokalisieren, wenn sie keinen AIS-Transponder an Bord haben³¹ oder das AIS außer Betrieb ist.

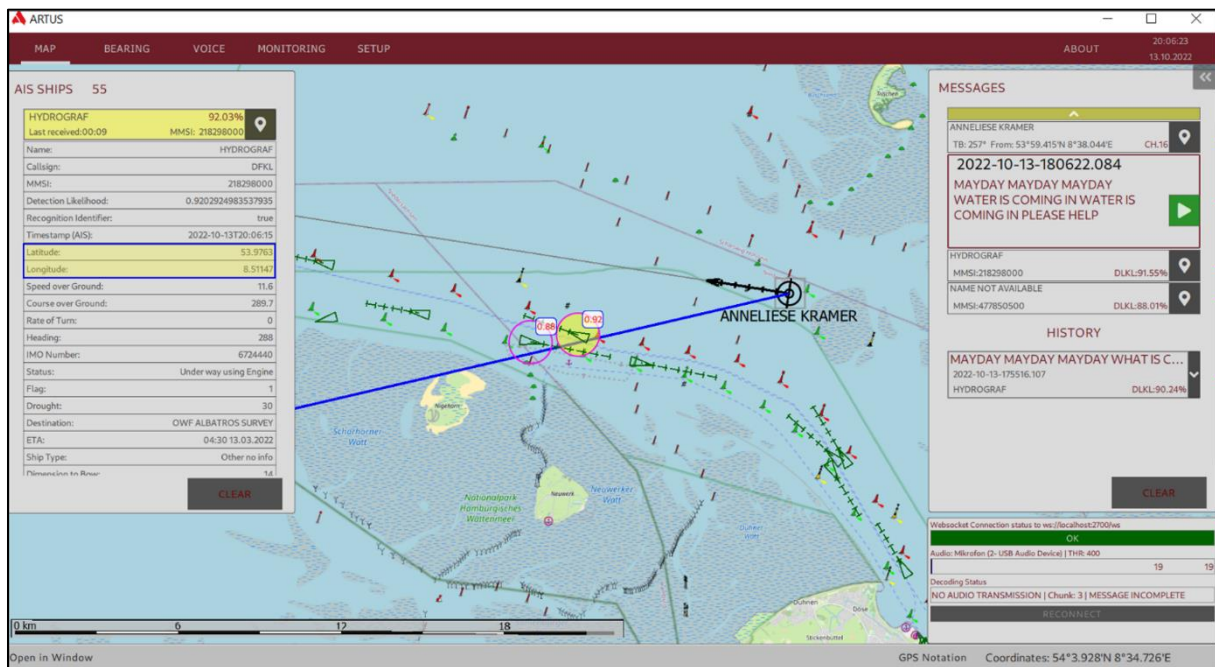


Abbildung 16 ARTUS-Benutzeroberfläche (Borddemonstrator)³²

In einer anschließenden Feedbackrunde hatten die Teilnehmer jeweils die Möglichkeit, ihre Eindrücke mit dem Projektkonsortium zu teilen. Hierbei ging es zum einen darum, die Umsetzung der im ersten Endnutzer-Workshop³³ ermittelten Anforderungen zu überprüfen. Zum anderen konnten Anregungen für Weiterentwicklungen und Transferideen gesammelt werden. Nach der Abfrage des Ersteindrucks wurden strukturiert die Bereiche Seekarte, Peilung und Lokalisierung sowie Transkriptanzeige, jeweils hinsichtlich ihres Informations- und Funktionsumfangs sowie der Darstellung auf der ARTUS-Benutzeroberfläche erfragt. Abschließend hatten die Endanwender die Möglichkeit, mögliche Vorschläge zur Ergänzung bzw. Verbesserung einzelner Funktionalitäten zu äußern und das Unterstützungspotenzial der ARTUS-Anwendung im eigenen Arbeitsalltag einzuschätzen.

Insgesamt zeigte sich eine durchweg positive Resonanz aller Teilnehmer und es wurde deutlich, dass alle Endnutzergruppen in der Zukunft von den Ergebnissen des Forschungsvorhabens profitieren können. Die Workshop-Teilnehmer waren vom technischen Reifegrad der präsentierten Lösung beeindruckt. Hier konnten vor allem die intuitive Bedienung der Software, die Leistungsfähigkeit des Spracherkenners und die Möglichkeiten der Peilervernetzung überzeugen. Als potenzielle Erweiterungen im Bereich der Visualisierung wurden u.a. eine farbliche Kennzeichnung der Transkriptanzeige in Ampelfarben, sowie die Möglichkeit zur Hervorhebung besonders dringlicher Meldungen in der Transkriptionshistorie angeregt. Bezüglich der Aufzeichnung der empfangenen Funksprüche wurden Mehrkanalfähigkeit bzw. Scan-Modus angesprochen. Weiterhin wurden intensiv die Möglichkeiten und Grenzen der Verwendung der

³⁰ Nachdem der zunächst gewählte zweidimensionale Ansatz (Kreuzung von zwei Geraden) nicht zu einem zufriedenstellenden Ergebnis hinsichtlich der Genauigkeit des Peilergebnisses führte, entschied sich RHOETHETA für einen dreidimensionalen Ansatz, in dem die Peillinien der Erdkrümmung entsprechend als sog. "Orthodrome" (=kürzeste Verbindung zweier Punkte auf einer Kugeloberfläche) in die Berechnung der Kreuzpeilung eingehen.

³¹ Dies ist regelmäßig bei kleineren Booten bzw. in der Sport- und Freizeitschiffahrt der Fall.

³² Quelle: RHOETHETA Elektronik GmbH

³³ Vgl. S. 4.

OpenSeaMap-Lösung im ARTUS-System im Vergleich zu einer Plug-in-Lösung in die bordeigene ECDIS sowie allgemeine Fragen der Hardwareintegration und Displaynutzung diskutiert. Insgesamt konnte das Projektkonsortium aus den Interviews von wertvollen Anregungen für die weitere Optimierung des ARTUS-Systems sowie hinsichtlich weiterer zukünftiger Anwendungsbereiche (u.a. im Ausbildung- und Prüfungsbereich) profitieren.

Die Endnutzer-Workshops wurden aufgezeichnet³⁴ und die Ergebnisse schriftlich dokumentiert und ausgewertet.

Wegen der weiterhin aufrechtzuerhaltenden Infektionsschutzmaßnahmen musste auch die Abschlussdemonstration am 31.05.2022 in dem in den Endnutzer-Workshops eingesetzten Online-Format durchgeführt werden. Auch hier wurden mit Rückgriff auf den speziellen Versuchsaufbau mit simulierten Datenströmen die Funktionen und die Leistungsfähigkeit des ARTUS-Systems erfolgreich präsentiert.

Ist es gelungen das ARTUS-System unter Laborbedingungen fortlaufend zu optimieren und erfolgreich zu testen, so steht die Erprobung im realen Einsatzkontext noch aus. Hierbei wird sich zeigen, inwieweit die entwickelte Lösung in der Lage ist, die besonderen Herausforderungen der in der Vorhabensbeschreibung identifizierten praktischen Anwendungsbereiche (auf See, in Küstennähe sowie in der Binnenschifffahrt) zu meistern. Hier werden ggf. noch Anpassungen und Optimierungen erforderlich werden; die bisherigen Erkenntnisse sowie die positive Resonanz der Endnutzer bilden aber bereits jetzt eine solide Grundlage für die zukünftige Nutzung der ARTUS-Technologie in der Einsatzpraxis. Die Projektpartner haben sich darauf verständigt, auch nach dem Ende der Projektlaufzeit die praktische Erprobung voranzutreiben, um den Weg der ARTUS-Technologie in die Anwendung und zur Unterstützung der Einsatzkräfte in der Seenotrettung weiter zu ebnen.

2.1.6. AP 6 Flankierende F&E-Aktivitäten

Im sechsten Arbeitspaket war die strategische Verwertung der Projektergebnisse vorzubereiten.

TAP 6.1 Einbindung in bestehende Lehrgänge

In diesem Teilarbeitspaket sollten bestehende Lehrgangsangebote auf eine Implementierbarkeit der ARTUS-Technologie betrachtet werden. Im Rahmen der Endnutzer-Workshops konnte noch einmal das große Potenzial für Ausbildungs- und Prüfungszwecke betätigt werden. Von Ausbilderseite wurde der Einsatz der ARTUS-Technologie insbesondere zur Unterstützung der Lehrgänge mit hohen Kommunikationsanteilen ausdrücklich begrüßt.

Gemeinsam mit dem CML wurden simulationsgestützte Lehrgänge und Trainings selektiert, die sich durch den Einsatz des ARTUS-Systems modifizieren bzw. verbessern lassen. Die Seenotretter-Akademie hat das ARTUS-System als wünschenswerte Funktionserweiterung für das SAR-Simulatorzentrum identifiziert und dessen Implementierung im Rahmen des im Berichtszeitraum gestarteten SIFO-Projektes AMARIS³⁵ vorgeschlagen. Außerdem hat ein erstes Gespräch mit dem Prüfungsamt Bremen stattgefunden, den ARTUS-Spracherkenner künftig als Unterstützungssystem in der Funkausbildung/-prüfung in Betracht zu ziehen.

TAP 6.2 Risikobewertung

In diesem TAP waren mögliche Risiken zu erfassen und zu bewerten, die mit der Einführung bzw. Verwendung der ARTUS-Technologie in der Einsatzpraxis auftreten könnten.

³⁴ Nach vorheriger Information und Zustimmung der Teilnehmer.

³⁵ Aeronautische und maritime Innovationsumgebung für interorganisationale Simulationen.

Wie bereits in den Ausführungen zu TAP 4.2 angesprochen,³⁶ konnten im Austausch mit Schiffsführern potenzielle Risiken im Bordbetrieb des ARTUS-Systems antizipiert werden. Die Ergebnisse konnten risikominimierend bei der Wahl des Installationsortes des Bord-Demonstrators auf der Schiffsbrücke des Seenotrettungskreuzers berücksichtigt werden.

Weitere potenzielle Risiken werden ggf. erst in der praktischen Erprobung sichtbar werden und werden im Rahmen weiterer Forschung & Entwicklung entsprechende Berücksichtigung finden.

TAP 6.4 Kommunikation/Publikation

Bedingt durch die COVID-19-Pandemie konnten die Forschungsergebnisse des Teilvorhabens insbesondere in der zweiten Projekthälfte nicht im vorgesehenen Ausmaß kommuniziert werden. Absagen von Konferenzen, allgemeine Reisebeschränkungen sowie spezielle Restriktionen bei der Durchführung von Dienstreisen behinderten die Möglichkeiten, die interessierte Fachöffentlichkeit zeitnah über Fortschritte und Teilergebnisse der Forschungsarbeiten zu informieren.

Unabhängig davon war die DGzRS an der Veröffentlichung eines Fachartikels (peer-reviewed) im Journal of Marine Science and Engineering beteiligt (Koordinierung und Konzeptionierung). Des Weiteren wurde das ARTUS-Projekt beim World Maritime Rescue Congress im Juni 2019 in Vancouver, BC (Vortrag und Poster, beides peer-reviewed), im Rahmen eines Webinars am University of California Berkeley Center for Catastrophic Risk Management (CCRM) im September 2019 sowie im Rahmen eines Vortrags beim Allgäuer Bergrettungstag im Dezember 2020 vorgestellt.³⁷

TAP 6.5 Datenschutz und rechtliche Aspekte

(Datenschutz-)Rechtliche Aspekte waren sowohl unter entwicklungs- als auch unter nutzungsbezogenen Aspekten zu berücksichtigen.

Für die Entwicklung des Spracherkenners wurde Trainingsmaterial u.a. aus Audiodaten generiert, die von BRR von Kanal 16 aufgezeichnet und dem Bereich Forschung & Entwicklung zur Verfügung gestellt wurden. Die Erfassung des Zeitstempels und die der Nachricht selbst zu entnehmenden Absenderinformationen (Nennung des Absendernamens bei Absetzen des Funkspruches) lassen grundsätzlich Rückschlüsse auf die Person des Absenders zu. Die Sprachaufzeichnung selbst enthält zudem personenbezogene Daten (Stimme) des Sprechers. Die Audiodaten wurden datenschutzkonform im Sinne der DSGVO gespeichert.³⁸

Bezogen auf die Nutzung der ARTUS-Software bleibt festzustellen, dass diese zum jetzigen Stand der Entwicklung ausschließlich auf das Projektkonsortium beschränkt ist. Weitere (lizenz-) rechtliche Fragen sind im Zuge der weiteren Verwertung neu zu bewerten und anzupassen. Dies betrifft auch Fragen der Speicherung der empfangenen Sprachmeldungen, also die Frage, wie lange aufgezeichnete Audios und Transkriptionen abgerufen werden können/müssen bzw. ob sie nach einer bestimmten Zeit automatisch gelöscht werden.

³⁶ Vgl. S. 26.

³⁷ Vgl. auch Punkt 2.6 auf S. 39.

³⁸ Vgl. Art 5 Abs. 1e) i.V.m. Erwägungsgrund 159 DSGVO (Verarbeitung zu wissenschaftlichen Forschungszwecken). Die physische Speicherung erfolgte bei der DGzRS auf einem separaten Rechner (offline System), der in einem abschließbaren Schrank aufbewahrt wurde. Die Daten wurden mit VeraCrypt verschlüsselt. Eine Weitergabe der Daten an den Unterauftragnehmer zum Training des Spracherkenners erfolgte ebenfalls verschlüsselt.

2.1.7. Zusammenfassung und Plan-/Ist Vergleich

Insgesamt konnten die wesentlichen Ziele des Projektvorhabens erreicht werden. Es liegt ein Assistenzsystem zur automatischen Transkription von UKW-Funkkommunikation und gleichzeitiger Absenderermittlung über Funkpeilung vor, das als Gesamtsystem im Laborversuch erfolgreich getestet werden konnte. Demonstratoren für den Land- und Bordbetrieb stehen zur Installation im Einsatzkontext bereit.

Pandemiebedingt konnte der Zeitplan des Forschungsvorhabens wegen der besonderen Schutzinteressen der involvierten Rettungsorganisationen nicht wie ursprünglich geplant eingehalten werden. Die Meilensteinziele konnten dennoch alle erreicht werden. Einzelne Verzögerungen wurden durch entsprechende Strategie- und Technologiewechsel sowie eine zwendungsneutrale Projektverlängerung um vier Monate aufgefangen. Zusammenfassend wurden folgende Maßnahmen ergriffen:

- Integration des Spracherkenners bereits im Early Demonstrator
- Einstellung einer weiteren Mitarbeiterin zur Generierung von Audiotranskripten
- Technologiewechsel beim Spracherkenner von *Kaldi* zu *Wav2Vec*
- Schaffung einer Simulationsumgebung zur Erprobung der Demonstratoren
- Umfassende Nutzung von Online-Formaten sowohl zur Durchführung von Arbeits- und Verbundtreffen als auch für die Durchführung von Endnutzer-Workshops und Live-Demonstrationen

Nach eingehender Prüfung sowie Abwägung von Kosten-Nutzen-Überlegungen und in Abstimmung der Verbundpartner mit den assoziierten Partnern wurden einzelne Teilziele modifiziert bzw. nicht weiter verfolgt³⁹. Der erfolgreichen Testung im Labor wird nach erfolgter Installation der Demonstratoren die Erprobung im realen Einsatzkontext folgen. Hierfür werden die Verbundpartner auch über das offizielle Projektende hinaus zusammenarbeiten. Auch bei den assoziierten Projektpartnern besteht großes Interesse an der weiteren Kooperation, insbesondere bezüglich der Nutzung der Spracherkenner-Komponente.

Abb. 17 gibt einen Überblick über den zeitlichen Ablauf des Projektes im Vergleich mit der ursprünglichen Planung.

³⁹ Dies betrifft im Wesentlichen die Anzahl der erstellten Demonstratoren.

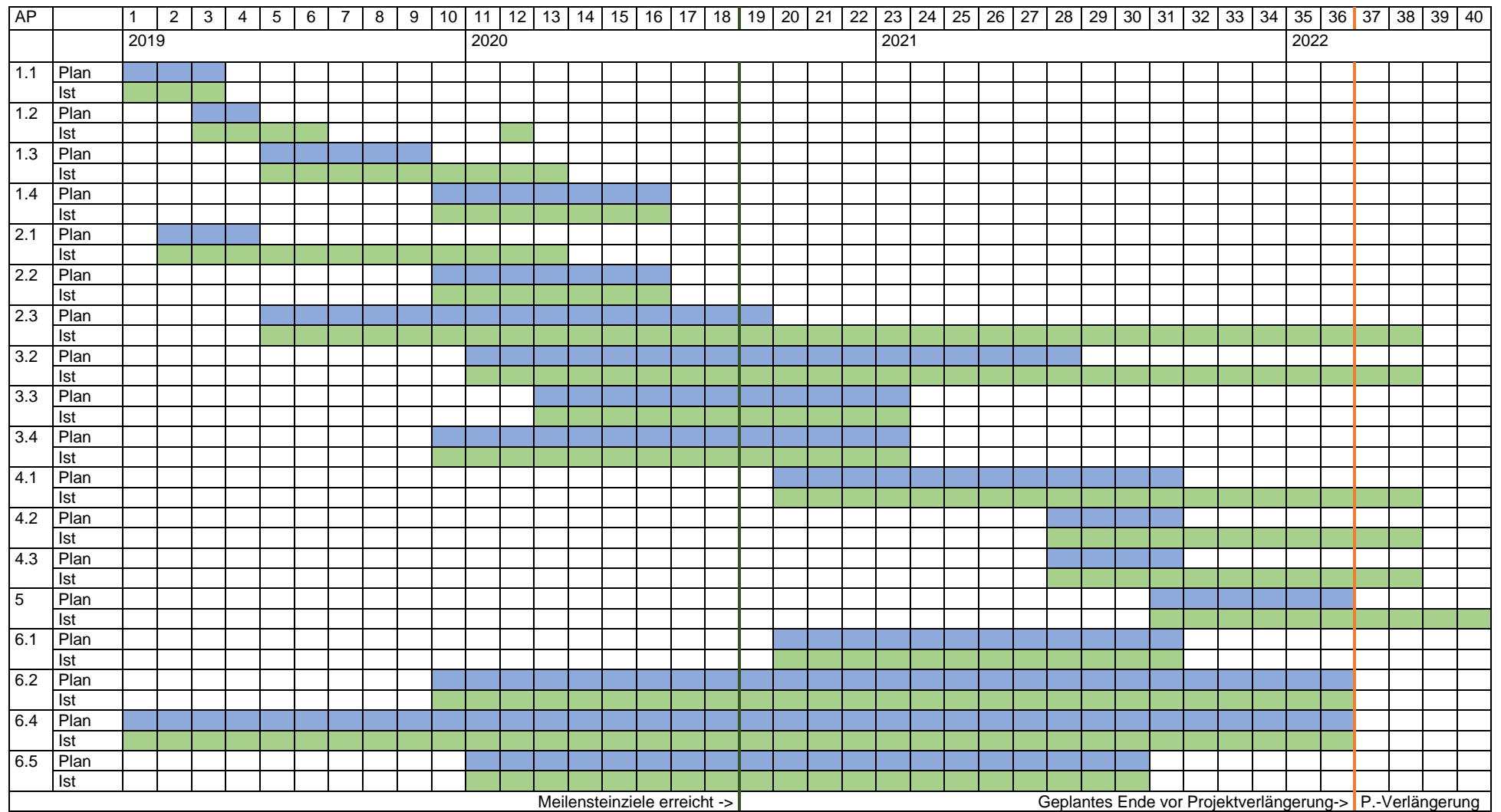


Abbildung 17 Plan-Ist Vergleich Zeitplanung

2.2. Wichtige Positionen des zahlenmäßigen Nachweises

Insgesamt sind die entstandenen Ausgaben des Vorhabens unter den im Gesamtfinanzierungsplan vorgesehenen geblieben. Einsparungen ergaben sich bei allen Ausgabenarten mit Ausnahme der Personalkosten. Die Mehrausgaben bei den Personalkosten (Position 0812) durch die zusätzliche Einstellung einer Wissenschaftlichen Mitarbeiterin⁴⁰ wurden durch die Einsparungen bei der Anzahl der gefertigten Demonstratoren (Position 0835), der pandemiebedingt stark reduzierten Reisekosten (Position 0846) sowie der von Microsoft kostenlos zur Verfügung gestellten MS-Teams-Plattform (Position 0834) mehr als aufgefangen.

2.3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Wie bereits dargestellt, konnte die Implementierung eines Erfassungssystems bei Bremen Rescue Radio aus verschiedenen Gründen nicht wie geplant stattfinden, so dass auf diese Weise kein zusätzliches Trainingsmaterial aus den UKW-Funkdaten von Kanal 16 gewonnen werden konnte.⁴¹ Um dennoch eine ausreichende Menge an domänenspezifischen Trainingsdaten für den Spracherkenner generieren zu können, wurden daher im Oktober 2020 die personellen Ressourcen für die manuelle Transkription von Audiodaten erhöht.

Zum einen konnte dadurch das Transkriptionsvolumen quantitativ zügig erhöht werden. Zum anderen führte das „4-Ohren-Prinzip“ auch zu einer Steigerung der Qualität des Trainingsmaterials, weil durch gegenseitige Kontrollen und Unterstützung bei der Bearbeitung schwer verständlicher Worte oder Passagen die Vollständigkeit der Transkriptionen insbesondere hinsichtlich der korrekten Erfassung der Schiffsnamen deutlich erhöht werden konnte.

Trotz der auf diese Weise sichergestellten kontinuierlichen Zuführung von Trainingsdaten stellte sich die Verbesserung der Spracherkennungsgüte zunächst nicht im gewünschten Ausmaß ein. Erst der vom Unterauftragnehmer angestoßene Technologiewechsel⁴² brachte einen deutlichen Fortschritt hinsichtlich der Reduzierung der WER.

Der zunächst beschrittene Weg (*Kaldi*-Technologie)⁴³ war unter den gegebenen Voraussetzungen folgerichtig und entsprach in der Konzeptions- und Anfangsphase des Projektes dem Stand der Technik. Der Bereich der Spracherkennung und der neuronalen Netze unterliegt jedoch stetigen Innovationen, deren Richtung und Ausmaß nur begrenzt antizipiert werden können. Neue Konzepte und Verfahren sind oft noch nicht ausreichend wissenschaftlich überprüft und so ist ein Technologiewechsel in einem laufenden Forschungsvorhaben immer auch mit einem gewissen Risiko behaftet. Im vorliegenden Projekt führte er jedoch zu einer deutlichen Verbesserung der Spracherkennungsqualität und so konnte das Projektkonsortium von der Flexibilität und Innovationsfreude des Unterauftragnehmers profitieren. Die Tatsache, dass die *Wav2Vec*-Technologie nur noch einen Bruchteil transkribierter Daten erfordert,⁴⁴ führte dazu, dass wenige Stunden zusätzlichen Materials auch gegen Ende der Projektlaufzeit noch einmal einen deutliche Qualitätsverbesserung bei der Spracherkenner-Komponente bedeutete.⁴⁵ Insofern erwies es sich als angemessen, das TAP 2.3 noch bis kurz vor dem offiziellen Projektende fortzuführen.

⁴⁰ Vgl. hierzu auch die Ausführungen zu TAP 2.3 sowie im nachfolgenden Abschnitt 2.3

⁴¹ Vgl. S. 16.

⁴² Vgl. S. 19.

⁴³ Vgl. S. 12 ff.

⁴⁴ Vgl. S. 21.

⁴⁵ So ermöglichte die Transkription von Audiodaten eines aktuellen und kompletten mehrstündigen realen Rettungseinsatzes mit insgesamt über 20 involvierten Sucheinheiten (Schiffe und Helikopter), der aufgrund der besonderen Einsatzgegebenheiten fast lückenlos über den UKW-Kanal 16 koordiniert wurde (andere Kanäle sowie Mobilfunk konnten wegen des Lage des Einsatzgebietes nicht genutzt werden), die kurzfristige Berücksichtigung von weiterem einsatzspezifischen Vokabular in deutscher und englischer Sprache in das Training der final implementierten Version 3.0 des Spracherkenners.

2.4. Nutzung der Ergebnisse und Fortschreibung des Verwertungsplans

Als ein für primär für die Unterstützung des SAR-Personals auf See entwickeltes Assistenzsystem eröffnen sich auch nach Projektende aussichtsreiche Möglichkeiten der Nutzung- und Weiterentwicklung der Forschungs- und Entwicklungsergebnisse.

Wie bereits dargelegt, hat die ARTUS-Technologie trotz der ausstehenden praktischen Erprobung im Einsatzkontext aus der Sicht der Endanwender bereits einen beachtlichen technologischen Reifegrad erreicht. Außerdem sehen die Endnutzer ein deutliches Unterstützungspotenzial in ihrem Arbeitsalltag und finden einen Großteil der zu Beginn des Projektes erhobenen Anforderungen im ARTUS-System umgesetzt. Zunächst werden die bereitstehenden Demonstratoren im jeweiligen Einsatzkontext zu implementieren und zu testen sein, um auf der Grundlage der zu gewinnenden Erkenntnisse das ARTUS-System weiter zu optimieren. Für eine zusätzliche Erhöhung des Technologiereifegrads werden – unabhängig von der ursprünglich im Rahmen des Projektes vorgesehenen Erprobung – noch weitere Forschungs- und Entwicklungsaufwendungen notwendig sein, um das System einsatznah noch gründlicher zu testen und den Demonstrator auf den Stand eines Einsatzdemonstrators und schließlich vom Prototyp bis zur Marktreife weiterzuentwickeln. Die DGzRS selbst wird die Ergebnisse selbst nicht wirtschaftlich verwerten, hat aber die Absicht das ARTUS-System innerhalb der nächsten Jahre flottenweit an Bord ihrer Einsatzmittel zu implementieren. Dabei ist vor dem Hintergrund der gewonnenen Erkenntnisse auch eine separate Weiterentwicklung und Verwertung der beiden Komponenten als sinnvolle Option in Erwägung zu ziehen.⁴⁶

Mit dem IAIS wurde eine Zusammenarbeit in der Weise vereinbart, als auch Trainingsmaterial in Form von transkribierten Audiodaten aus Rettungseinsätzen, das über das Projektende hinaus seitens der DGzRS geliefert wird, für die Weiterentwicklung der Spracherkenner-Software genutzt werden soll. RHOTHETA als Hardwarehersteller plant eine Vertiefung der Zusammenarbeit mit der Firma *JAKOTA Cruise Systems* und ist an einer wirtschaftlichen Verwertung der Forschungsergebnisse interessiert.

Hinsichtlich der inhaltlichen Weiterentwicklung des ARTUS-Systems, wie etwa die domänen-spezifische Adaptierung der Spracherkennungs-Komponente kann der Anwendungsbereich *auf See* - trotz der herausragenden Bedeutung der englischen Sprache in der maritimen Kommunikation – auf weitere Sprachen ausgedehnt werden.⁴⁷ Wie gezeigt wurde, erfordert die *Wav2Vec*-Technologie einen deutlich reduzierten Bedarf an annotierten Daten für das Training der Spracherkenner-Komponente, so dass eine Erweiterung über die englische und deutsche Sprache hinaus ohne größeren Aufwand möglich sein wird.

Dies eröffnet eine große Bandbreite an Anwendungen in diversen Einsatzszenarien und Zielbranchen. So könnte der ARTUS-Spracherkenner beispielsweise ebenfalls für die Unterstützung der Einsatzkoordination in Leitstellen und Einsatzmitteln von Rettungsdiensten *an Land* und *in der Luft* angepasst werden.

Über die reine Nutzung der Spracherkenner-Komponente hinaus, wäre ein Einsatz der ARTUS-Technologie auch in der Luftfahrt denkbar. Anstelle der Verwendung von AIS wäre die Implementierung der Senderlokalisierungskomponente mittels ADS-B (Automatic Dependent Surveillance Broadcast)-Transpondern denkbar.

Ein weiteres Einsatzfeld läge im Ausbildungs- und Prüfungsbetrieb der Funkausbildung. Mit Hilfe des Spracherkenners wäre es möglich, die Ausbildung effizienter zu gestalten und eine (zusätzliche) Dokumentationsmöglichkeit für den praktischen Prüfungsbetrieb zu schaffen. Das gewonnene Trainingsmaterial (Audiodaten und Transkripte) bietet darüber hinaus

⁴⁶ Vgl. auch S. 26.

⁴⁷ Vgl. hierzu auch Fußnote 14 auf S. 15.

vielfältige Möglichkeiten hinsichtlich der Beurteilung der Seefunkkommunikation. Die DGzRS hat bereits damit begonnen, interne Fallstudien zur Evaluation der Einsatzkommunikation durchzuführen.

Nach dem Projektende kann zusammenfassend ein nicht unbeträchtliches Verwertungspotenzial der ARTUS-Technologie konstatiert werden.

Abschließend bleibt noch zu erwähnen, dass eine seitens des CML – ohne Berücksichtigung des schöpferischen Anteils der DGzRS – während der Bewilligungsphase des Projektes eingereichte Patentanmeldung an der ARTUS-Technologie durch Recherchen der DGzRS in den Veröffentlichungen des DPMA bekanntgeworden ist. Die DGzRS hat aus Ressourcengründen auf einen Rechtsstreit mit der Fraunhofer-Gesellschaft verzichtet, ohne dass es zu einer Einigung gekommen ist.

2.5. Bekanntgewordene Ergebnisse Dritter

Im Berichtszeitraum sind keine unmittelbar vorhabensrelevanten Ergebnisse von dritter Seite bekannt geworden; die Synthese von KI-basierter Spracherkennung und intelligenter Peilung kann somit als Alleinstellungsmerkmal des Forschungsvorhabens betrachtet werden.

Gleichwohl waren im Berichtszeitraum einige Projekte auszumachen, die sich in verschiedenen Anwendungskontexten mit Spracherkennung auf der Grundlage neuronaler Netze beschäftigt haben. Darüber hinaus unterliegt der Bereich der automatisierten Spracherkennung einem schnellen Wandel, so dass es wichtig war, die Entwicklungen in diesem Bereich genau im Blick zu behalten.

Unmittelbaren Einfluss auf die Leistungsfähigkeit der Spracherkenner-Komponente des vorliegenden Forschungsvorhabens hatten die Arbeiten von *Schneider et al.* (Schneider, Baevski, Collobert, & Auli, 2019) zum unüberwachten Lernen und *Baevski et al.* (Baevski, Zhou, Mohamed, & Auli, 2020) zum selbstüberwachten Lernen auf Audiodaten. Diese zeigten, dass der Einsatz der *Wav2Vec*-Technologie mit signifikant geringeren Mengen an annotierten Daten zu gleichen bzw. besseren Ergebnissen im Vergleich zu den State of the Art, d.h. *Kaldi*-basierten Spracherkennungsalgorithmen führen, welche erheblich größere Trainingskorpora erfordern.

Aus anwendungsorientierter Perspektive ist im Berichtszeitraum das Projekt NotAs (Unterstützung der Notrufannahme durch KI-basierte Sprachverarbeitung), unter der Leitung des Instituts für Feuerwehr und Rettungstechnologie Dortmund zu nennen.⁴⁸ Die DGzRS war hier als assoziierter Partner involviert. Ziel des Vorhabens war es, einen digitalen, multilingualen⁴⁹ Sprachassistenten zu entwickeln, der das Einsatzpersonal in Leitstellen bei der Annahme von Notrufen in ihnen nicht geläufigen Sprachen unterstützen soll. Der Sprachassistent sollte nicht nur eine korrekte Übersetzung, sondern durch die Implementierung einer Dialogfunktion auch eine zielgerichtete Gesprächsführung ermöglichen. Durch eine automatisierte Interpretation des Sprachflusses sollte das System darüber hinaus in der Lage sein, einsatzrelevante Informationen zu extrahieren (Namen, Adressen, etc.) und diese als Ausfüllhilfe in das zu erstellende Einsatzformular zu übertragen. Aus datenschutzrechtlichen Gründen konnten für das Training des Spracherkenners wegen der enthaltenen sensiblen personenbezogenen Informationen keine Originalnotrufe verwendet werden. Daher wurden entsprechende simulierte Notrufgespräche in deutscher, englischer und polnischer Sprache aufgenommen, anschließend manuell transkribiert und durch Übersetzungen ergänzt. Das entstandene Material wurde zum Training eines auf Standardkommunikation vortrainierten, kommerziellen

⁴⁸ Vgl. https://www.sifo.de/sifo/shareddocs/Downloads/files/projektumriss_notas.pdf?__blob=publicationFile&v=2

⁴⁹ Im Rahmen des Forschungsvorhabens wurden die Sprachen Deutsch, Englisch und Polnisch einbezogen.

Spracherkenners genutzt. Dadurch konnten die WERs im Projektverlauf zwar reduziert werden, jedoch war zum Projektende trotz erster erzielter Erfolge noch deutlicher Verbesserungsbedarf bei der Spracherkennungskomponente festzustellen (Pratler-Wanczura & Aschenbrenner, 2022), (Haack, 2022). Das ARTUS-Vorhaben war hier insgesamt, u.a. auch aufgrund der Nutzung eines bereits speziell vortrainierten Spracherkenners, durch den Zugang zu umfangreichem, domänenspezifischem Originalaudiodatenmaterial aber auch aufgrund des vollzogenen Technologiewechsels deutlich im Vorteil.

Als weiteres Forschungsvorhaben unter deutscher Leitung, das im Berichtszeitraum startete und sich mit der Nutzung Künstlicher Intelligenz für das Notfallmanagement in Krisensituationen befasst, ist das Projekt SPELL (= Semantische Plattform zur intelligenten Entscheidungs- und Einsatzunterstützung in Leitstellen und beim Lagemanagement) zu nennen.⁵⁰ Unter der Leitung des DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) soll eine KI-basierte, hybride Plattform entwickelt werden, die helfen soll, das Einsatzpersonal in Rettungsleitstellen bei der Beherrschung der Informationsflut in Krisensituationen zu unterstützen. Einsatzrelevante Daten aus verschiedensten Quellen sollen in einer neuen Art und Weise aufbereitet und bereitgestellt werden. Die intelligente Spracherkennung ist dabei jedoch nur eine Komponente und es bleibt abzuwarten, inwieweit es gelingen wird, eine leistungsfähige Spracherkennungstechnologie bereitzustellen, die in der Lage ist, kontextbezogen zur Entscheidungsunterstützung in kritischen Einsatzsituation beizutragen.⁵¹

Auf internationaler Ebene konnten im maritimen Bereich zwei Vorhaben ausgemacht werden, die sich im Berichtszeitraum mit Assistenzsystemen im maritimen Seefunk beschäftigt haben.

Zum einen ist das Projekt der Swedish Maritime Administration (SMA) zu nennen, die gemeinsam mit den beiden schwedischen IT Unternehmen *Tentifty AB* und *Maranics AB* ein Forschungs- und Entwicklungsprojekt zur Schaffung eines KI-basierten Notruf-Erkennungssystems („emergency call detection system“) initiiert hat, um im JRCC (Joint Rescue Co-ordination Center) gezielt Notrufmeldungen zu erkennen und zu transkribieren.⁵² Der Fokus liegt dabei auf der Identifizierung von notruftypischen Schlüsselwörtern, wie z.B. „Mayday“, „SOS“, „help“, „sinking“ etc. Das System wird auf Basis der im JRCC eingehenden Funkmeldungen trainiert. Sobald das System eine Notrufmeldung identifiziert, wird dies dem Operator auf dem Display zusätzlich hervorgehoben übermittelt. Außerdem besteht die Möglichkeit, die Nachricht erneut abzuspielen und manuell Nutzerkorrekturen vorzunehmen⁵³. Auf diese Weise wird das System kontinuierlich und auf realen Daten trainiert (Youd, 2021). Insofern wurde hier ein dem ursprünglich auch bei Bremen Rescue Radio geplanten vergleichbares Erfassungssystem realisiert. Dennoch bleiben einige Fragen offen, so ist nicht ersichtlich, welche Technologie verwendet wurde, wie leistungsfähig und zuverlässig das System arbeitet (WER), und ob es auch in der Lage ist, multilingual zu transkribieren. Außerdem ist unklar, inwieweit das cloudbasierte System auch außerhalb des JRCC, also an Bord von Einsatzmitteln zuverlässig funktionieren kann.

Als weiteres europäisches Projekt mit dem Fokus Seefunkkommunikation ist das italienisch-französische Interreg Projekt ISIDE (Innovazione per la Sicurezza di Mare) zu nennen.⁵⁴ Dieses verfolgte unter der Leitung der Universität Cagliari in Zusammenarbeit mit sieben weiteren

⁵⁰ <https://www.dfki.de/web/news/projekt-spell-ki-plattform-rettungsleitstellen>; Laufzeit: 01.06.2021-31.05.2024

⁵¹ Vgl. hierzu erste Ergebnisse in einem auf dem 8th Workshop on Formal and Cognitive Reasoning am 19. September 2022 in Trier vorgestellten Paper (Maletzki, Rietzke, & Bergmann, 2022), in dem ein Entscheidungsunterstützungssystem („Copilot“) zur Bearbeitung medizinischer Notrufe auf der Basis von transkribierten Anruferaufzeichnungen unter Verwendung neuronaler Netze vorgestellt wurde.

⁵² Vgl. <https://www.sjofartsverket.se/sv/om-oss/nyheter-och-press/nyheter/sjofartsverket-anvander-artificiell-intelligens-inom-sjo--och-flygraddningen/>

⁵³ In welchem Ausmaß eine solche Feedbackkontrolle aber im Einsatzkontext vor dem Hintergrund der Arbeitsbelastung tatsächlich durchgeführt werden konnte, wurde nicht offengelegt.

⁵⁴ <https://interreg-maritime.eu/web/iside/>

Partnern (u.a. Küstenwache, Hafenbehörden) das Ziel, über die Verarbeitung und Übermittlung von Textnachrichten ein Messaging-System zur Reduzierung von Missverständnissen in der Schiff-Schiff- bzw. Schiff- Land-Kommunikation zu schaffen. Solche schriftlichen Nachrichten können - als Ergänzung zur verbalen Kommunikation über Funk - mittels AIS-Geräten entweder als vordefinierte SMCP-Standardkurzmeldungen oder als Freitextphrasen („safety related messages“) verarbeitet und übertragen werden. Das zu entwickelnde Konzept soll als benutzerfreundliche Schnittstelle auf AIS-Anlagen zum Einsatz kommen und lässt sich vereinfacht als Assistenzsystem zur Übersetzung schriftlich eingehender Nachrichten in die Arbeitssprache des Bordpersonals⁵⁵ beschreiben. Ebenso ist es möglich, (voreingestellte) Textnachrichten mit sofortiger Übersetzung zu verschicken. Es werden auch Antworten auf gestellte Fragen vorgeschlagen und über Drop-Down-Menüs können weitere Informationen (z.B. Schiffsname, Position, etc.) übertragen werden. Für Nutzer ohne AIS ist auch eine Anbindung über DSC (Digital Selective Call) oder Internet möglich, die Kopplung an ein mobiles Endgerät (Laptop, Tablet) erfolgt über eine spezielle Software/Hardware-Schnittstelle. Das System soll primär in Gebieten mit hoher Verkehrsdichte zum Einsatz kommen,⁵⁶ um die Risiken, die sich aus einer fehlerhaften oder mehrdeutigen Kommunikation in kritischen oder Notfallsituationen ergeben können, zu minimieren.

Mit dem ARTUS Projekt lässt sich das beschriebene Vorhaben insofern nicht vergleichen, als es sich hier um ein Assistenzsystem für die schriftliche Kommunikation handelt. Es bietet keinerlei Spracherkennungsfunktion für verbale Funkkommunikation und bleibt insofern weit hinter den Möglichkeiten der ARTUS-Technologie zurück.

Abschließend sollen noch einige internationale Projekte aus der Luftfahrt betrachtet werden, die sich im Berichtszeitraum mit Fragen der KI-basierten Spracherkennung beschäftigt haben. Auf der Applied Human Factors and Ergonomics Conference (AHFE) 2019 in Washington, DC präsentierte die U.S. Naval Air Warfare Center Training Systems Division ein militärisches F&E-Projekt zu "Automatic Speech Recognition (ASR)" zur Nutzung in der Kommunikation zwischen Kampffjet-Piloten und den Bodenkontrollstationen. Der Projektleiter *John Killilea* berichtete, dass trotz des hochgradig standardisierten Vokabulars in der militärischen Luftfahrt, der von ihnen genutzte Spracherkenner erhebliche Probleme hat, muttersprachliche Kommunikation von Piloten aus dem mittleren Westen der USA zu erkennen. Zur Behebung dieses Problems waren 80 Stunden domänenspezifisches Trainingsmaterial erforderlich. Der finale Spracherkenner soll lediglich in den Bodenkontrollstationen zum Einsatz kommen.

Auf der 21. INTERSPEECH Conference in Shanghai im Oktober 2020 stellte ein internationales Forschungskonsortium das European Union-Horizon 2020 Projekt CleanSky EC-H2020 ATCO2 vor (Zuluaga-Gomez, Motlicek, Zhan, Vesely, & Braun, 2022).⁵⁷ Dieses Vorhaben zielt auf die Entwicklung einer Spracherkennungsplattform, für welche die Kommunikation der Bodenkontrolle gesammelt, organisiert und automatisch vorverarbeitet werden sollte. Trainiert wurde auf mehr als 170 Stunden ATC (Air Traffic Control)-Sprachdaten. Neben der reinen Sprachkommunikation wurden auch kontextuelle Informationen in Form von Meta-Daten berücksichtigt. Im Rahmen des Vorhabens wurden verschiedene Fragestellungen untersucht. Von besonderem Interesse waren hier die sich ändernden Akzente/das unterschiedliche Vokabular auf verschiedenen Flughäfen.⁵⁸ Es konnte gezeigt werden, dass die akzentübergreifenden Fehler aufgrund der Menge der Daten minimiert werden konnten. Das entwickelte Spracherkennungs-System zeigt eine durchschnittliche WER von 7,75 % über vier Datenbanken. Bei der Analyse der WER konzentrierte man sich jedoch zunächst nur auf die korrekte

⁵⁵ Hier wurde primär die Übersetzung von Englisch ins Italienische bzw. umgekehrt betrachtet.

⁵⁶ Es wurde in den Hafengebieten von Genua, Livorno, Cagliari, Toulon und Bastia getestet.

⁵⁷ Vgl. auch <https://www.atco2.org>; ATCO = Automatic Collection and Processing of Voice Data from Air-Traffic Communication

⁵⁸ Hier waren Deutsch, Tschechisch, Französisch, Griechisch, Italienisch, Schweizerisch sowie amerikanisches Englisch relevant.

Erkennung der Callsigns (Kocour, et al., 2021). Aufgrund der hohen Sicherheitsanforderungen im Luftverkehr kann ein solches System nicht „on the fly“, d.h. in der realen Betriebsumgebung, entwickelt und getestet werden. Es muss also vor dem ersten Praxiseinsatz bereits einen sehr hohen Entwicklungsstand nachweisen können.

Das DLR-Institut für Flugführung leitete im Berichtszeitraum das EU Projekt HAAWAI (Highly Automated Air-Traffic Controller Workstations with Artificial Intelligence Integration) und widmete sich unter Beteiligung von Flugsicherungen aus Österreich, Island, Großbritannien und Kroatien der Fragestellung, wie Fluglotsen mittels KI entlastet werden können und gleichzeitig die Sicherheit im Luftverkehr verbessert werden kann. Ziel war es, ein Entscheidungsunterstützungssystem zu entwickeln, das in der Lage ist, das Verhalten von Fluglotsen vorherzusagen. Im Fokus standen auch hier die Berücksichtigung verschiedener Akzente sowie Abweichungen von der Standardterminologie und die eingeschränkte Qualität der Sprachsignale.⁵⁹ Während frühere Projekte zur Spracherkennung im Luftverkehr⁶⁰ vor allem darauf zielten, eine automatisierte Pflege der Fluginformationen zu unterstützen, die Fluglotsen also von der manuellen Eingabe der Fluglotsenbefehle in das Flugverkehrsmanagement-System zu entlasten, wurde im HAAWAI-Vorhaben als eine zentrale Messgröße für die Leistungsfähigkeit der Spracherkennung, die Erkennung von fehlerhaften Funkbestätigungen durch Piloten (sog. "readback errors") betrachtet. Neben der Callsign-Erkennungsrate sollte hier damit zusätzlich auch die Befehlserkennungsrate optimiert werden (Harfmann & Teodor, 2021).

Mit Blick auf das ARTUS-Vorhaben bleibt anzumerken, dass die Übertragbarkeit der Forschungsergebnisse aus dem Luftverkehr in den maritimen Einsatz wegen der hohen Spezifität (Besonderheiten in Vokabular und Akzenten) limitiert ist.⁶¹ Zusammenfassend kann festgestellt werden, dass auch jüngere Forschungsvorhaben zur Spracherkennung im aeronautischen Kontext ausschließlich auf der *Kaldi*-Technologie basieren. Ein nutzbringender Einsatz im maritimen Kontext würde daher erhebliches zusätzliches domänenspezifisches Trainingsmaterial erfordern. Erst nach Abschluss des ARTUS-Vorhabens wurde die erste Arbeit überhaupt öffentlich, die das Potenzial der *Wav2Vec*-Technologie im Bereich der Spracherkennung in der Luftverkehrskommunikation untersucht (Zuluaga-Gomez, et al., 2022).

2.6. Geplante und realisierte Veröffentlichungen

Publikation

Gözzalan, Aylin; John, Ole; Lübcke, Thomas; Maier, Andreas; Reimann, Maximilian; Richter, Jan-Gerrit; Zverev, Ivan: Assisting Maritime Search and Rescue (SAR) Personnel with AI-Based Speech Recognition and Smart Direction Finding. In: Journal of Marine Science and Engineering, 20. Oktober 2020. (peer-reviewed)

Präsentationen

World Maritime Rescue Congress im Juni 2019 in Vancouver, BC (Vortrag und Poster, beides peer-reviewed)

Webinar am University of California Berkeley Center for Catastrophic Risk Management (CCRM) im September 2019

Vortrag beim Allgäuer Bergrettungstag im Dezember 2020.

⁵⁹ Vgl. <https://www.haawaii.de/wp>

⁶⁰ Z.B. das vom DLR geleitete EU-Projekt MALORCA (Machine Learning of Speech Recognition for Controller Assistance), vgl. <https://malorca-project.de>

⁶¹ Das HAAWAI-Vorhaben etwa bezog sich auf den Streckenflugverkehr in Island und den Luftverkehr im Nahbereich von London.

3. Pressespiegel

<https://www.hellenicshippingnews.com/communication-in-maritime-emergencies-artus-improves-vhf-radio-communication/>

https://www.seenotretter.de/presse/archiv/ansicht?tx_news_pi1%5Baction%5D=detail&tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Bnews%5D=1923&cHash=007044e13bf3928f41d1245c38232679

<https://www.cml.fraunhofer.de/de/forschungsprojekte1/ARTUS.html>

https://www.rettungsdienst.brk.de/fileadmin/Eigene_Bilder_und_Videos/LGST/Sicherheitsforschung/Abstract_Forschungsvorhaben_ARTUS__2_.pdf

4. Literaturverzeichnis

- Baevski, A. S., Fuegen, C., & Auli, M. (9. September 2019). *Wav2vec: State-of-the-art speech recognition through self-supervision*. Abgerufen am 2. Februar 2023 von <https://ai.facebook.com/blog/wav2vec-state-of-the-art-speech-recognition-through-self-supervision/>
- Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (22. Oktober 2020). *wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*. Abgerufen am 2. Februar 2023 von <https://arxiv.org/pdf/2006.11477.pdf>
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by jointly learning to align and translate. *ICLR 2015*, (S. 1-15). Abgerufen am 1. Februar 2023 von <https://arxiv.org/pdf/1409.0473.pdf>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (24. Mai 2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Abgerufen am 1. Februar 2023 von <https://arxiv.org/abs/1810.04805>
- Github*. (kein Datum). Abgerufen am 3. Februar 2023 von <https://github.com/cristoper/wav2vec>
- Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (25. Juni 2006). *Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks*. Abgerufen am 3. Februar 2023 von <https://dl.acm.org/doi/10.1145/1143844.1143891>
- Gref, M. (2022). *Robust Speech Recognition via Adaption for German Oral History Reviews*. Bonn: Diss.
- Haack, M. (29. September 2022). Erschließung neuer Technologien für das Absetzen von Notfallmeldungen,. Ulm. Abgerufen am 5. Februar 2023 von <https://m.youtube.com/watch?v=5rrClelOQ8U>
- Hannun, A. (2017). Sequence Modeling with CTC. A visual guide to Connectionist Temporal Classification, an algorithm used to train deep neural networks in speech recognition, handwriting recognition and other sequence problems. *Distill*. doi:10.23915/distill.00008
- Harfmann, J., & Teodor, S. (6. Juni 2021). *Speech Recognition in the ATC environment for Human Performance Evaluation, Radar Label Maintenance, Readback Error Detection*. Abgerufen am 10. Februar 2023 von 1st Stakeholder Workshop: https://www.haawaii.de/wp/wp-content/uploads/2021/07/03_2021-06-28-JulesTeo-HAAWAI-Applications.pdf
- Hunt, M. (1990). Figures of Merit for assisting connected-word recognisers. *Speech Communication*, 9(4), 329 ff.
- Jaitly, N. (2017). *Natural Language Processing with Deep Learning: Lecture 12: End-to-End model for Speech Processing*. (S. University, Hrsg.) Abgerufen am 25. Januar 2023 von <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/lectures/cs224n-2017-lecture12.pdf>
- Kocour, M., Vesely, K., Blatt, A. Z.-G., Cernocky, J., Klakow, D., & Motlicek, P. (2021). *Boosting of contextual information in ASR for air-traffic call-sign recognition*. Abgerufen am 9. Februar 2023 von https://publications.idiap.ch/downloads/papers/2021/Kocour_INTERSPEECH2021_2021.pdf
- Maletzki, C., Rietzke, & Bergmann, R. (9. September 2022). *Utilizing Expert Knowledge to Support - Medical Emergency Call Handling*. Abgerufen am 9. Februar 2023 von <https://ceur-ws.org/Vol-3242/paper6.pdf>

- Otter, D. W., Medina, J. R., & Kalita, J. K. (21. Dezember 2019). *A Survey of Deep Learning for Natural Language Processing*. *IEEE Transactions on Neural Networks and Learning Systems*. Abgerufen am 29. Januar 2023 von <https://arxiv.org/pdf/1807.10854.pdf>
- Perez, C. (23. Mai 2019). *The Paradigm Shift of Self-Supervised Learning*. Abgerufen am 2. Februar 2023 von <https://medium.com/intuitionmachine/the-paradigm-shift-of-self-supervised-learning-744a6819ce08>
- Povey, D., Ghoshal, A., & Boulianne, G. e. (2011). *The Kaldi Speech Recognition Toolkit*. Abgerufen am 25. Januar 2023 von https://publications.idiap.ch/downloads/papers/2012/Povey_ASRU2011_2011.pdf
- Pratler-Wanczura, S., & Aschenbrenner, D. (14. November 2022). *CRISIS PREVENTION - Fachportal für Gefahrenabwehr, Innere Sicherheit und Katastrophenhilfe. Braucht Feuerwehr Forschung?* Abgerufen am 5. Februar 2023 von <https://crisis-prevention.de/feuerwehr/braucht-feuerwehr-forschung.html>
- Scherer, A. (1997). *Neuronale Netze: Grundlagen und Anwendungen*. Braunschweig, Wiesbaden.
- Schneider, S., Baevski, A., Collobert, R., & Auli, M. (11. September 2019). *wav2vec: Unsupervised Pre-training for Speech Recognition*. Abgerufen am 1. Februar 2023 von <https://arxiv.org/pdf/1904.05862.pdf>
- Schulz, H., & Behnke, S. (2012). Deep Learning. *Künstliche Intelligenz*, 26, 357-363.
- Vaswani, A. S., & Parmar, N. e. (6. Dezember 2017). *Attention is all you need*. Abgerufen am 27. Januar 2023 von <https://doi.org/10.48550/arXiv.1706.03762>
- Youd, F. (17. August 2021). *Mayday Mayday! Answering emergency calls with help from AI*. Abgerufen am 9. Februar 2023 von <https://www.ship-technology.com/analysis/mayday-mayday-answering-emergency-calls-with-help-from-ai/>
- Zuluaga-Gomez, J., Motlicek, P., Zhan, Q., Vesely, K., & Braun, R. (13. August 2022). *Automatic Speech Recognition Benchmark for Air-Traffic Communications*. Abgerufen am 9. Februar 2023 von <https://arxiv.org/abs/2006.10304>
- Zuluaga-Gomez, J., Prasad, A., Nigmatulina, I.; Motlicek, P., Kleinert, M., Helmke, H., Zhan, Q. (22. Oktober 2022). *How does pre-trained wav2vec 2.0 perform on Domain-shifted ASR? An Extensive Benchmark on Air Traffic Control Communications, präsentiert auf dem 2022 IEEE Spoken Language Technology Workshop scheduled for 19-22 January 2023 in Doha, Qatar,.* Abgerufen am 9. Februar 2023 von <https://arxiv.org/pdf/2203.16822.pdf>