YRKESHÖGSKOLAN
NOVIA

Performance Assessment of a marine-adapted speech recognition engine

Ahmed Elhadi

Degree Thesis

Bachelor of Maritime Management

Degree Programme in Maritime Management, Captain

Turku, 2023

DEGREE THESIS

Author: Ahmed Elhadi
Degree Programme and place of study: Degree Programme in Maritime Management
Specialization: Bachelor of Maritime Management
Supervisor(s): Peter Björkroth, Katarina Sandström

Title: Performance Assessment of a marine-adapted speech recognition engine

---

Date: 20.4.2023 Number of pages: 46 Appendices: 1

Abstract

Lack of efficient marine communication poses a significant risk to the operational safety of the vessels, as communication problems are the cause of more than forty percent of accidents at sea. The introduction of speech recognition engines to the maritime domain carries a potential increase in communication efficiency by reducing misunderstandings on board and allowing for the effective handling of multiple conversations for shore-based operators.

The aim of this paper is to compare the transcription outcome, accuracy, and word error rate of a marine-adapted automatic speech recognition engine and a standard one. Consequently, assessing the worthiness of a specialized marine speech recognition library in developing marine VHF radio communication transcription engines. As a result, paving the way for decision-support and decision-making tools for shore-based Remote Operation Centers in the future.

Real VHF audio recordings were obtained and then processed to build a specialized marine speech recognition data library from scratch. The library is then used to fine-tune a general-purpose speech recognition engine by a third party, Lingsoft Oy. Consequently, the adapted engine and the standard 'control' engine are tested using sample data.

The results of the experiment showed an increase in the performance of the marine-adapted engine as compared to the standard engine in transcription outcome, accuracy, and word error rate. Further research is needed to explore the full potential of this method of adaptation, the volume and quality of the maritime data library, and the possible applications in the industry.

Language: English
Key Words: Communication, SMCP, AI, Speech recognition, Fine-tuning

Contents

1 Introduction

The maritime industry is a vital sector of the global economy, as it is responsible for transporting over 90% of the world's goods (International Chamber of Shipping, 2021), marking it as the backbone of global trade and the supply chain. According to the United Nations Conference on Trade and Development, the world fleet was more than 98,140 merchant vessels of 100 gross tons and above, with an estimated total trade volume of 11.08 billion tonnes in 2019 (United Nations Conference on Trade and Development, 2020). Moreover, nearly 2 million seafarers are serving on board vessels as of 2021 (International Chamber of Shipping, 2021).

The industry faces unique challenges related to safety, security, and environmental sustainability, and it continues to evolve as new technologies and regulations emerge. Particularly, safety is a top priority in the maritime industry, and strict regulations are in place to ensure the safe and efficient operation of vessels and ports. As Mrs Panagiota Chrysanthi emphasized on the safety challenges in shipping and the relevancy of the human factor during her speech at the Hellenic American Maritime Forum in Athens in 2019 (Chrysanthi, 2019).

Consequently, my thesis focuses on a small aspect of maritime safety that is, maritime communication. In general, maritime communication can be divided into three different streams: ship-to-shore, ship-to-ship and on-board communication. VHF radio is the most widely used method of marine communication. However, it has a lot of challenges such as background noise, voice fluctuation and distortion, strong accents, language barriers, and the fact that the unique maritime VHF radio terminology ’Standard Maritime Communication Phrases’ is used to a limited extent which directly jeopardizes the safety of the crew and the respective operations such as navigation, cargo handling, mooring and so on (Noble, 2007).

Moreover, maritime communication falls under a safety factor called human error. This broad category of human error is responsible for 75%-96% of marine casualties (Allianz Global Corporate & Specialty, 2012). Subsequently, marine communication complications and the misinterpretation of information attribute to a staggering one-third of the above-mentioned human error marine casualties (Ziarati, Ziarati, Bigland, & Acar, 2011).

Although the maritime industry is rapidly moving towards autonomous, digital and sustainable solutions, there is still room for improvement, and the scope of the maritime communication sector is not an exception. Hence comes the introduction of speech

recognition technology, which is a type of artificial intelligence that enables computers to interpret and understand human speech, and present it in a written format (IBM, n.d.). Such an initiative has the potential to revolutionize marine communication by enabling real-time, accurate, and secure communication between crew members on board, pilots, and shore personnel.

Speech recognition technology is already introduced and implemented in the maritime field within ambitious projects such as Automated Transcription of Maritime VHF Radio Communication for Search and Rescue 'SAR' Mission Coordination project 'ARTUS' where Fraunhofer CML is a partner (Fraunhofer CML, n.d.) and ELNAV, the Croatian start-up developing Helm order monitor, which provides a new approach to introducing speech recognition supporting systems to the bridge (ELNAV Advanced Safety Systems, n.d.).

In context, the market potential of speech recognition technology in the maritime domain is promising (Interview with Stormbom, 6 March 2023). There is a big market for the technology in the maritime industry (Interview with Nakilcioglu 5 April 2023). In numbers, the estimated annual revenues of utilizing this technology in the Maritime Rescue Coordination Centres MRCC and Vessel Traffic Services VTS is about 9 million euros, given that there are over 40 VTS centres in Europe alone. Moreover, this value goes up to 125 million euros if this technology is introduced on board ships and to other land stations. Globally, the worldwide potential of the market is 300 million euros in annual revenues. (Interview with Nakilcioglu 5 April 2023).

My thesis aims to explore the potential use of speech recognition technology in the maritime industry. Then consequently propose a new approach to fine-tune or adapt a general-purpose speech recognition model to the marine domain using a specialized maritime data library created for this experiment, and comparatively analyse its performance. Fine-tuning of speech recognition models refers to the process of modifying a pre-trained speech recognition model to improve its accuracy and performance on a specific task or domain.

2 Purpose and Problem Statement

With regards to the challenges and shortcomings within VHF radio communication, the need to introduce more digital solutions into the maritime domain, and the development of AI and speech recognition models, this paper introduces an approach to fine-tuning speech

recognition engines to the maritime industry using a specialized maritime data library created for the purpose of this experiment.

The aim of this paper is to assess the performance of a speech recognition engine after adapting it with a specialized maritime speech recognition data library, so as to assess the worthiness of introducing a specialized maritime speech recognition library. This is done using comparative analysis i.e. by comparing the transcription outcome, accuracy and word error rate of a marine-trained speech recognition engine and a standard one.

To achieve this aim a new specialized maritime speech recognition library is created from scratch. This research paves the way for the development of VHF radio communication decision-support and decision-making tools for shore-based Remote Operation Centers in the future.

3 Previous research

The literature review of this paper highlights some of the current applications of speech recognition engines in the maritime domain, the challenges of transcribing and interpreting maritime VHF radio recordings, along with previous research and experiments where the approach of fine-tuning the speech recognition models is utilized and an improvement is observed in the engine performance.

To start with, Fraunhofer CML is developing a deep neural network AI speech recognition engine as a part of their contribution to the ARTUS project (Fraunhofer CML, n.d.). The ARTUS project is an ambitious initiative aiming to facilitate the coordination and functioning of rescue operations at sea by automatically transcribing VHF radio transmissions into written form and the possible identification and localization of respective vessels in distress (Reimann & John, 2020). It acts as a communication-enhancing tool and a documentation feature.

The possibility of viewing the radio messages in a written form improves the ship-to-shore and ship-to-ship communication as it allows the radio operator to review the information exchanged with the other party in real-time or at a later stage (Interview with Nakilcioglu 5 April 2023; Reimann & John, 2020), hence reducing the risk of misunderstands or missing important operational details as it is a more effective way of representing information (Porathe, Eklund, & Göransson, 2021).

This benefits the maritime rescue coordination centres ‘MRCC’ by providing the radio messages in a document or transcripts form, hence reducing the time needed to investigate and collect relevant information on the rescue mission prior to initiating the actual operations (Reimann & John, 2020). Thus, saving time in a domain where a minute could cost a lot both economically and in the human factor.

One of the obstacles of this project is the challenges of VHF radio maritime communication, which is the most common mode of communication within the industry. VHF radio transmissions present a challenge to speech recognition engines and even to human operators due to their nature and acoustic conditions. Characteristics such as loud background noise, distortion, quality, and strength of the signal contribute to the difficulty of message interpretation for human operators and transcription for speech recognition engines. (Interview with Nakilcioglu 5 April 2023; Reimann & John, 2020).

Furthermore, when the radio operator is placed under the scope, a different set of challenges arise. Poor English language skills, the abundance of dialects and accents, low volume, fast speech, and code-switching, where the radio operator switches between two languages during the speech, all contribute to the unique nature of maritime VHF radio (Reimann & John, 2020).

Not to mention the domain-specific features of the maritime radio messages, consisting of the unique terminology and the simple grammar recommended by the IMO Standard Maritime Communication Phrases protocol (Reimann & John, 2020), which is unfortunately, not used as extensively and frequently by mariners and radio operators (Schriever, 2009). These challenges complement that it is less stressful for mariners to receive text-based navigational information compared to answering radio voice calls (Porathe, Eklund, & Göransson, 2021).

Another utilization of speech recognition technology in the maritime industry is with the ELNAV, which is a Croatian startup aiming to improve the safety of navigation by utilizing the latest technology (ELNAV, n.d.). Their product, Helm order monitor, uses AI to detect and monitor the correct execution of Helm orders. It utilizes various ship sensors to detect vessel movements, and speech recognition technology, which is developed by Fraunhofer IDMT (Fraunhofer IDMT, 2022), to interpret helm orders and compare them with the side where the helmsman turns the wheel. It acts as a lane departure warning system for cars.

This approach of identifying and tackling individual challenges in distinct areas on the ship such as the bridge and consequently building a suitable digital solution for it is moving into the industry. It aligns with the vision of the digitalisation of the maritime industry and slowly paves the way for the 3rd and 4th degrees of autonomy on Maritime Autonomous Surface Ships 'MASS' based on the regulatory framework created by the IMO (International Maritime Organization (IMO), 2021).

Continuing another aspect, fine-tuning of speech recognition engines to start with refers to the process of adapting a pre-trained model to perform better on a specific task or dataset. This involves training the model on a smaller, task-specific dataset, which is the specialized maritime library, that is related to the target domain, in addition to the large, general-purpose dataset on which the original model was trained. By fine-tuning the model on a more specific dataset, it can learn to recognize domain-specific language and context more accurately, leading to better overall performance. (Barreto, n.d.).

The fine-tuning of transformer-based neural networks can be employed to develop multilingual text and speech models at the time being. This approach has made it feasible to create tools that support several languages, even when there is a small amount of annotated data available for these languages (Guillaume, et al., 2022; Partanen, Hämäläinen, & Klooster, 2020; Prud'hommeaux, Jimerson, Hatcher, & Michelson, 2021). This also applies to domain-specific terminology and annotated data, which is the scope of this experiment. These transformer-based architecture models facilitate the process of updating them, as there is no need to rebuild or train the models again from scratch every time, which is the case with other algorithms (Interview with Nakilcioglu 5 April 2023).

This approach of fine-tuning speech recognition models is provided worthy and is deemed to be the way how people are, and will work with AI now, and in the future (Interview with Stormbom, 6 March 2023). Individuals and companies can download and use the general-purpose models that are pre-trained with a huge amount of data rather than training models from scratch. This process of pre-training speech recognition models is costly and is considered very demanding in terms of both processing power and finances. Hence not all companies are capable of such tasks. (Interview with Stormbom, 6 March 2023).

Experimentally, this method is utilized in (Conneau, Baevski, Collobert, Mohamed, & Auli, 2020) when the team fine-tuned their cross-lingual pre-trained model and tested its performance against baselines from previous work. Cross-lingual learning aims to build models which leverage data from other languages to improve performance. Their model

XLSR, when compared to the best results available, reduced the relative phoneme error rate by 72% on the CommonVoice benchmark. (Conneau, Baevski, Collobert, Mohamed, & Auli, 2020).

The phoneme error rate is a measure of the accuracy of a speech recognition system, which calculates the percentage of phoneme errors made by the system in transcribing speech. It is computed by dividing the total number of phoneme errors, such as incorrect or missed phonemes, by the total number of phonemes in the reference transcription. The lower the phoneme error rate, the more accurate the system is at recognizing speech. (Kurimo, 1997).

For clarification, the Common Voice benchmark is a publicly available evaluation framework that measures the performance of automatic speech recognition 'ASR' models on a standardized dataset of speech recordings from the Common Voice project. The benchmark is designed to help researchers and developers compare the accuracy of different ASR systems and track progress over time. (Conneau, Baevski, Collobert, Mohamed, & Auli, 2020; CommonVoice Mozilla, n.d.)

Furthermore, in (Guillaume, et al., 2022), the approach used is similar to this paper. First, a multilingual general-purpose speech recognition engine was acquired. A data library with audio recordings and time stamps is then created using manual transcription to teach the system how to match the incoming audio or speech with the text labels. The fine-tuning or adaptation is finally done using the data library created. The findings indicated an improvement in the quality of phonemic transcriptions compared to previous experiments. (Guillaume, et al., 2022).

On another topic, the post-processing of speech recognition output is also a technique to reduce errors and improve the engine's results. Post-processing of speech recognition output refers to the process of analysing the transcribed text generated by a speech recognition system and correcting any errors or improving its readability. Post-processing is typically performed using a combination of rule-based and machine learning-based methods and can be automated or done manually by a human editor. (Pekichev, 2021).

This approach of post-processing the output of speech recognition engines to reduce errors and increase accuracy is not novel, with some of its benefits highlighted by Eric K. Ringger as early as 1996 (Ringger & Allen, 1996). In his paper, a new post-processing technique is introduced to reduce transcription errors by refining the vocabulary, or the character-data

library, of the speech recognition model. A similar approach that also aligns with the views of this paper (Ringger & Allen, 1996).

Moreover, in (Anantaram & Kopparapu, 2017), the aim is also to develop domain-specific speech recognition engines by re-purposing general-purpose speech recognition ones. This approach is taken to prevent building a new interface, as a new interface must be built for domain-specific engines if a new domain is introduced, or an existing one is updated. This also aligns with the motives of this paper by utilizing fine-tuning technologies to minimize the workload and avoid training speech recognition models from scratch, which is considered a worthy approach (Interview with Nakilcioglu 5 April 2023; Interview with Stormbom, 6 March 2023).

Their paper introduces two methods of post-processing of ASR output, Eco-Devo, and machine learning, prior to further NLP or Natural language processing. Results of the experiments and an assessment of the adaptation methods' worthiness are carried out by comparatively analysing the accuracy of the 4 different engines introduced in the experiment. The results of the experiments show that the 2 mechanisms are "*promising*". (Anantaram & Kopparapu, 2017, p. 18).

In context, Natural language processing or NLP is a subfield of artificial intelligence or AI that focuses on the interaction between computers and human language. NLP enables computers to understand, interpret, and generate human language in a way that is meaningful and useful (Avasthi, 2021). In the realm of language technology, NLP and voice recognition are distinct yet interconnected fields. Voice recognition is the transcription of spoken language into text, whereas NLP is the processing of the text to understand and interpret its meaning. While voice recognition can function independently of NLP, the opposite is not true since NLP cannot directly analyse audio input. (Picovoice.ai, 2022).

With regards to the scope of this paper, the data library created for this experiment features two parts, the manual transcriptions along with the corresponding time stamps forming the data labels, and the conversion of some transcriptions to the Standard Maritime Communication phrases 'SMCP'. The first phase, the data labels, is used to fine-tune the speech recognition model. Consequently, the SMCP dataset created does not contribute to the fine-tuning, and can only be utilized in the post-processing phase or NLP at a later stage (Interview with Nakilcioglu 5 April 2023).

It is to be noted that no papers or experiments are found that illustrate the process of building a specialized maritime speech recognition data from scratch. Moreover, no previous work on the fine-tuning of speech recognition models to function specifically in the maritime domain is found.

3.1 SMCP and Maritime English

This chapter highlights the significance of the human factor in marine casualties, especially in the communication aspect. Moreover, it illustrates the initiatives proposed by the IMO to reduce accidents, before finally discussing the relevancy of the Standard maritime communication phrases ’SMCP’ and citing a few sources on the fact that SMCP is used to a limited extent in practice.

The majority of marine casualties can be traced back to 'human error,' a wide-ranging category that is believed to account for anywhere between 75%-96% of such incidents. This type of error can be linked to various factors, but two of the most commonly cited are competition pressures, often emanating from shore-based entities, and fatigue. This is mostly evident in busy shipping areas where crewmembers and shore personnel may have limited opportunities to rest. (Allianz Global Corporate & Specialty, 2012).

In another study, a percentage of 80% is found to be the contribution of ’human error’ to marine accidents (Verbeck, 2011). Out of this staggering percentage of human-induced casualties, about one-third are attributed to communication problems and misinterpreting of information. (Ziarati, Ziarati, Bigland, & Acar, 2011; Ziarati, Safety At Sea – Applying Pareto Analysis, 2006).

The accident that occurred in 1996 involved the crude oil carrier Sea Empress, which spilled a significant amount of oil into the sea off the coast of Milford Haven in the U.K. Unfortunately, language barriers made the situation worse, as a nearby large Chinese tugboat could not be utilized to assist due to communication difficulties. (Perez & Manuel, 2003). Similarly, during the Scandinavian Star accident in 1990, language barriers likely contributed to the loss of life as a result of inter-ship communication difficulties (Perez & Manuel, 2003; Porathe, Eklund, & Göransson, 2021).

To reduce these accidents, English was established as the operational language of international shipping. Consequently, the IMO introduced the competence ”*Use the IMO Standard Marine Communication Phrases and use English in written and oral form*” (IMO

STCW Manila 2010, Table A-II/1) in the STCW. This competence falls under Table A-II/1 titled ’Specification of minimum standard of competence for officers in charge of a navigational watch on ships of 500 gross tonnage or more (International Maritime Organization IMO, 2010).

Suprisingly, maritime communication in all its forms be it ship to ship, ship to shore or on-board communication has a staggering percentage of 90% of non-native English speakers (Pritchard, 2003). Mariners are expected to have satisfactory English language skills, proper domain-specific terminology ’SMCP’, and good communication skills (International Maritime Organization IMO, 2010; Ahmmed, 2017).

Then the Standard maritime communication phrases ’SMCP’ protocol was introduced in November 2001, which replaced the Standard Marine Navigational Vocabulary ’SMNV’ adopted by the IMO in 1977 (International Maritime Organization (IMO), n.d.) and SEASPEAK, a technical and standardised form of English for a specific purpose, task and context. (Porathe, Eklund, & Göransson, 2021).

The Standard Maritime Communication Phrases ’SMCP’ is a domain-specific dictionary, or a set of standardized English phrases used in the maritime industry for effective and clear communication between ships and shore-based personnel, covering various topics such as navigation, safety, and general operations (Porathe, Eklund, & Göransson, 2021). The aim of the SMCP is to provide a standard for maritime communication among multi-lingual radio operators, crews and maritime personnel so as to avoid misunderstandings and increase the efficiency of the maritime communication (Dževerdanović-Pejović, 2013). Hence increasing the safety of navigation, the crew and other ship operations (Reimann & John, 2020).

However, despite the introduction of SMCP as a competence in the STCW, and the widespread efforts to train and familiarize mariners with it, SMCP and Maritime English is still used to a limited extent (Noble, 2007; Kataria, 2011). The challenges or shortcomings of maritime communication were not solved with the adoption of the SMCP protocol (Reimann & John, 2020; Dževerdanović-Pejović, 2013).

It is confirmed that actual maritime radio VHF communication recordings ”*do not display high SMCP content*” (Noble, 2007, Abstract; Schriever, 2009; Kataria, 2011), especially in ”*situations of immediate danger or heavy traffic*” (Dževerdanović-Pejović, 2013, IJTTE, p. 394). It consists of so-called causal English speech with nautical terminology (Dževerdanović-Pejović, 2013; Kataria, 2011). Also, local languages are abundant with

small barges, and it may be used too when communicating with local mooring launches, tugs, pilots and so on. (Kataria, 2011).

In (Schriever, 2009), questionnaires were received from 132 seafarers from 17 nationalities asking 32 questions. The question to be concerned with is *"How often was the IMO publication 'Standard Maritime Communication Phrases' or SMCP used at sea?"* (Schriever, 2009, IMEC 21, p. 56). The findings were interesting. Almost two-thirds or 64.2% of the respondents stated that the publication was never, or rarely used by them. The figure below illustrates their results. This issue is seen to be partly because of the teaching bodies and their training programs. (Schriever, 2009).

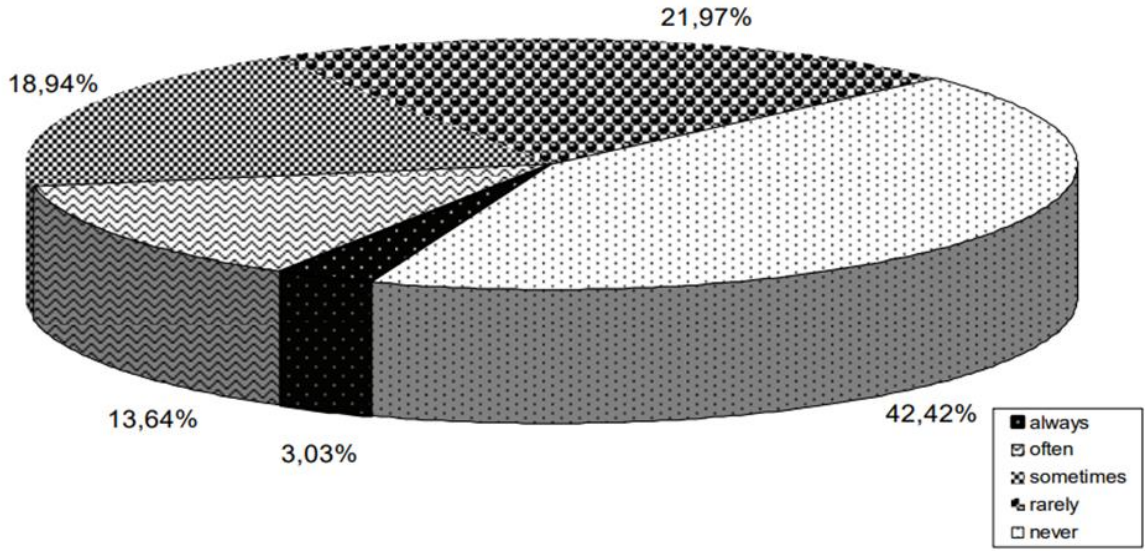

Figure 1. The frequency with which the SMCP was used on board (Schriever, 2009)

3.2 Introducing AI

Apart from the normal ways to improve communication in the maritime industry, such as training courses, refreshers, and forcing bills. Technology can be introduced. Such as with the ELNAV start-up use case (ELNAV Advanced Safety Systems, n.d.) mentioned on page 4.

As an end goal, the availability of a reliable marine communication speech recognition model will provide a written transcript of maritime radio communication that will allow for more effective communication and fewer errors and repetitions. Although, the SMCP and training arguments from Chapter 3.1 make it challenging to train a system to understand and interpret marine radio operators.

In the maritime domain, no tool that automatically transcribes VHF radio messages is currently available for commercial use and hence it is not possible to rely on text-based information exchange (Reimann & John, 2020; Interview with Stormbom, 6 March 2023), or arguably even text-based information display for confirmation and bookkeeping both on board and onshore. Advances in AI, computer processing power, specialized language models and deep neural networks exhibit a potential for the introduction of such supportive technology to the maritime industry in the future.

On the shoreside, such systems would enable overloaded radio operators to manage multiple conversations easier (Interview with Nakilcioglu 5 April 2023) by having, for instance, a visionary UI where each vessel's conversation transcripts and their respective data are displayed on the screen. In addition to having a smart radio direction finder and AIS data available, the operator would be provided with a more organized, informative, and user-friendly UI that acts as a decision-support tool. Also, in the event of a successful interpretation of human speech, all sorts of text analysis can be done on the transcripts such as a trigger system feature to get the attention of the operator in case of urgency or distress for instance, which also acts as a valuable decision-support tool (Interview with Stormbom, 6 March 2023).

Such tools would help shore-based radio operators with tracking and keeping up with multiple conversations at the same time, and multi-tasking (Interview with Nakilcioglu 5 April 2023; Reimann & John, 2020). Also, having radio communication in written form or transcripts facilitates the work of rescue operations and distress calls. The documentation of radio messages can facilitate investigations on marine accidents as well. (Reimann & John, 2020). Commercially, and in practice, Fraunhofer CML is developing a deep neural network AI speech recognition engine as a part of their contribution to the ARTUS project as mentioned on page 3.

This approach can also be implemented on board different types of ships such as commercial cargo vessels, pilot boats, coast guard boats, fishing boats, yachts and so on. A screen on the bridge that transcribes what the navigational personnel hears on the radio is deemed beneficial in many ways (Interview with Nakilcioglu 5 April 2023; Porathe, Eklund, & Göransson, 2021). It is a more effective way of representing information as it reduces misunderstandings, lowers the possibility of missing information, and allows the operator to rely less on his/her memory (Porathe, Eklund, & Göransson, 2021). Not to mention that such

a product is initially scalable (Interview with Nakilcioglu 5 April 2023) and can act as an operational safety supplement tool.

From a futuristic view, the speech recognition models and algorithms may also be enhanced to interpret, analyse, and translate different languages to English. Minimizing possible misunderstandings, and language limitations with ship-to-ship and ship-to-shore communication. Such an initiative would require multi-lingual maritime-language-specific data to be created.

Finally, and as discussed in Chapter 3 under natural language processing or NLP, the possible NLP applications for a maritime dataset allows the machine to understand and interpret maritime communication and consequently develop a wide variety of solutions such as communication assessment features in Intelligent Learning systems 'ILS' in simulators to train students and mariners for instance.

4 Research Methodology

The aim of this paper is to assess the performance of a speech recognition engine after adapting it with specialized maritime speech recognition data. This is done using comparative analysis i.e. by comparing the transcription outcome, accuracy and WER of a marine-trained speech recognition engine and a standard one.

Unfortunately, and to start with, a specialized maritime speech recognition data library does not exist in the market, or it was not deemed possible to find or acquire. Hence, a new sample library is built for this experiment.

The library is in the form of transcribed text and SMCP text along with its respective audio files. Real-life VHF radio communication audio files are acquired and transcribed from voice to text using Subtitle edit, an open-source subtitle generator, editor, and video subtitle tool (Niske, n.d.). This tool is recommended by a third party, Lingsoft Oy (Lingsoft Oy, n.d.), which is the technical partner in charge of fine-tuning the speech recognition engine with the specialized maritime data library.

Then, sections of the transcribed text are converted to Standard Marine Communication Phrases using a basic Excel sheet, prior to a thorough studying and understanding of the IMO SMCP Resolution A.918(22), adopted on the 29th of November 2001 (IMO, 2001).

Two speech recognition engines are used for testing procedures. One engine is adapted to the maritime domain using the newly built specialized maritime speech recognition library. It is referred to as a ‘Specialized maritime speech recognition engine’ or ’Adapted engine’. This engine is adapted and fine-tuned by the third party, Lingsoft Oy (Lingsoft Oy, n.d.). Information regarding the fine-tuning process is acquired through an interview with Lingsoft personnel in the Appendix. Hence, given the uniqueness of the engine, there are no alternatives to it available.

The second engine is not adapted and is used as a control for the experiment. It is referred to as ‘Standard speech recognition engine’ or ’Standard engine’. A basic online search is done, and eventually, some transcription software are found and dug into more. Table 1 represents the transcription tools found along with some comments on their capabilities. For clarification, the criterion for the standard engine is:

1. Automatic speech recognition tool
2. AI-powered features
3. Not trained with any maritime data
4. Marketed as being able to transcribe any audio without limitations

Table 1. Transcription tools and comments

Transcription tool	Comments
Otter (Otter.ai, n.d.)	Real-time transcribing only, cannot transcribe audio recordings
FTW Transcriber (the FTW Transcriber, n.d.)	No AI features stated
Temi (Temi, n.d.)	Difficult audios results are 'Mostly unusable' as stated on their website. Also a paid service.
Happyscribe (Happyscribe, n.d.)	Manual transcription tool. An automatic engine is needed
oTranscribe (oTranscribe, n.d.)	Manual transcription tool. An automatic engine is needed
Descript (Descript, n.d.)	Fits all criteria of the standard engine

Hence, The engine model used for this task is Descript. It is an innovative video editor with AI-powered features such as automatic transcription, video editing, podcasting, and clip creation (Descript, n.d.). Descript is chosen as it is the only tool found that fits the criteria for the standard engine.

Then, sample data is prepared and run through both engines and results are obtained. Sample data consists of four different-themed audio recordings. First, pilot audio recording which involved pilot boarding operations communication. Second, a broadcast file with weather forecasts, warnings, coast guard messages, and urgency ‘sécurité transmissions. Finally, two random sets of audio data from the United States, and Australia. The sample data is a total of 20 minutes of maritime VHF radio recordings.

Results are in the form of text transcripts. Comparative analysis is then carried out. Initial transcripts produced by the Marine expert in the first stage of the experiment are used as reference data for the comparison.

Three factors are used to assess the results. Transcription outcome, transcription accuracy, and word error rate, WER. Transcription outcome refers to the characters, or words identified and hence transcribed by the engine. Transcription accuracy determines how accurate the characters generated by the speech recognition engine are. Word error rate, WER is a comprehensive measure that combines both the outcome and accuracy.

Another factor, the outcome speed, which refers to how fast the engine generates the results, is not taken into consideration so as to not deviate from the research goals. No other relevant factors are identified and hence the experiment proceeds with the three factors mentioned above.

The transcription outcome is deduced using a basic character counter to find the number of words generated by the engine. An online tool called WORDCounter is used (WORDCounter, n.d.). Many tools are available online and provide the same quality of service and hence the above-mentioned tool is just chosen. On the other hand, the process of deducing the transcription accuracy involved a few steps. First, a text similarity software is used to obtain the initial similarity percentage using the Marine expert’s transcripts as the reference data.

The engine transcripts of the test samples are run through Toolsaday (TOOLSADAY, n.d.), an AI-based text analysis and similarity tool. The purpose of this step is to find the base

percentage of similarity between the two texts, again, using the marine expert results as a reference point, and assuming they are 100% correct at this step.

Then, the number format error is corrected. This error is due to the standard speech recognition engine transcribing the numbers as digits, and hence it cannot be used for comparison. The first step is to extract all the numbers from the transcripts' text using Browserling number extractor online tool (broswerling, n.d.). Then, the number of digits extracted was counted using WORDCounter (WORDCounter, n.d.).

The last step is to calculate the number of *Accurate digits*, i.e. The numbers transcribed by the engine accurately, by multiplying the *Number of digits* from the previous step with the respective sample's *initial 'similarity percentage* from Toolsaday. For instance, 70% of the 41 digits extracted correspond to 29 accurate digits for X sample file.

Next, the same is done for the *Total engine character count* from the previous transcription outcome stage. The *Total engine character count* generated by the respective engines is multiplied by the *initial similarity percentage*. Then their product is added to the *Accurate digits* to find the *Total matching characters,* which is the number of characters accurately transcribed by the engines. Then, this *Total matching characters* is compared with the *Total engine character count* from the transcription outcome stage to find the corrected transcription accuracy percentage. Finally, the Marine expert's accuracy is assumed to be 95%. So, the *Corrected similarity percentage* is to be interpolated to the 95% and then the weighted accuracy percentages are calculated to find the final transcription accuracy of each engine.

The word error rate, by definition, is a measure used to evaluate the performance of automatic speech recognition or text-to-speech systems. It represents the percentage of words in the transcribed output that differ from the words in the reference transcript. This includes substitutions, deletions, and insertions from the reference transcript. The lower the WER, the better the performance of the system. (Fox, 2021). The introduction of WER as a reliable metric is proposed in (Interview with Nakilcioglu 5 April 2023).

Two distinct approaches are used to find the WER, the first is an online free-to-use WER calculator tool called Amberscript (Amberscript, n.d.). Based on the research made, no other tools are found that provide this service, and hence the experiment is proceeding with Amberscript.

The second approach involved some preparations with the engine transcripts and used high-level programming language Python (Python Software Foundation, n.d.). The Python package JiWER is installed, and the right code is run to compare the respective engines' transcripts to the reference transcripts generated by the Marine expert (Python Package Index PyPI, 2023). The use of the JiWER package was only possible after installing and adding Poetry to the computer's path. Poetry is a "*dependency management and packaging*" tool in Python (Python Poetry, n.d.).

Regarding the transcripts, the specialized-maritime engine transcripts do not require any modifications. On the other hand, the Marine expert and the Standard engine transcripts are misaligned and required the lines and paragraphs to be merged and combined into a single paragraph for each sample file. VBA, or Microsoft Visual Basic for Applications of Microsoft Office Word is used where the code "*Merge multiple lines into one single paragraph*" is compiled and executed (ExtendOffice, n.d.).

Moreover, the Standard engine transcripts had the digits in a numerical form. These digits are converted to textual form. For instance, 62 is converted to sixty-two throughout the transcripts. This is done to ensure the same formatting is applied to all transcripts and to increase the reliability of the WER results.

5 Specialized Maritime speech recognition engine

This chapter provides a comprehensive walkthrough on the building process of the specialized maritime speech recognition engine. Initially with the maritime data creation, and later with the technical adaptation. Chapters 5.1 and 5.2, or Phases I and II respectively, explain in depth the process of building the specialized maritime library from scratch. Consequently, Chapter 5.3, or Phase III, illustrates how Lingsoft Oy (Lingsoft Oy, n.d.) utilizes the newly-built maritime data library to adapt the speech recognition model.

5.1 Phase I: Transcription

Real-life VHF radio communication from 5 different regions is acquired. The US, Canada, the UK, Australia, and Europe. The audio files were not recorded in any simulated or controlled environment, and they entitle actual maritime VHF radio transmissions of various conditions such as weather forecasts, coast guard warnings, ship-to-ship communication, and pilotage operations to name a few.

The audio files are not simulated as the goal is to prepare a set of data that considers both the interference of radio signals, which can be simulated, and the strong accents, dialects, and the 'actual' operational communication taking place in practice, with all its limitations.

5.1.1 Transcription process

The software used for the transcription process is called Subtitle edit. It is an open-source subtitle generator, editor, and video subtitle tool. (Niske, n.d.). The data generated is saved in the form of .srt file format.

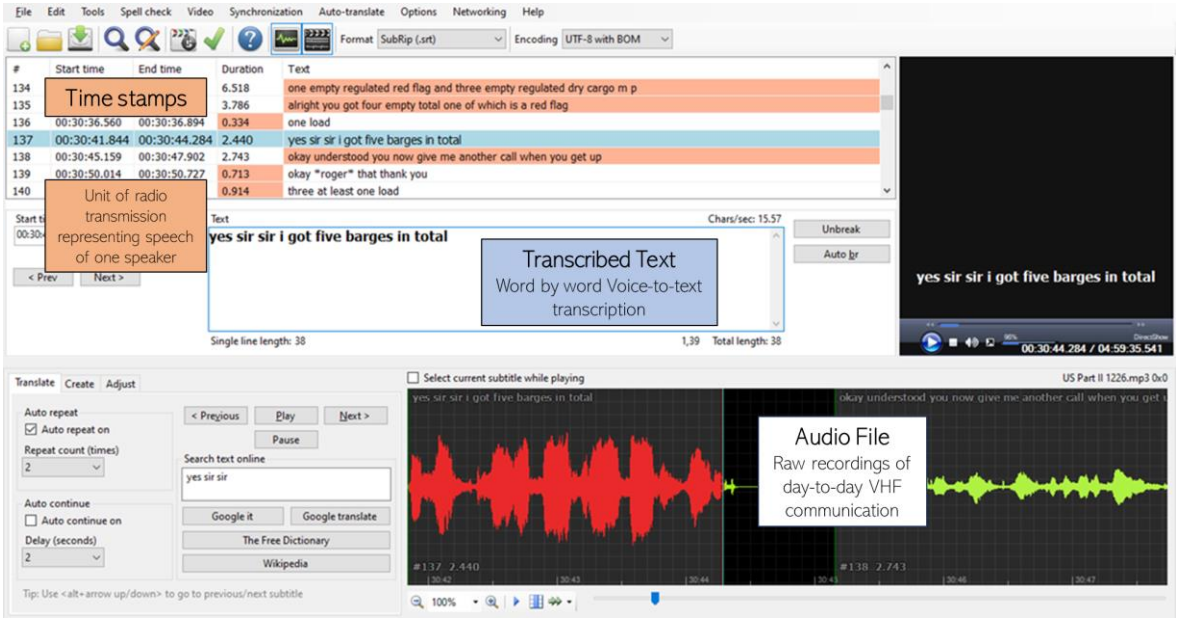

Figure 2. Transcription process software and practicality

As shown in Figure 2 above, the transcription process is based on time stamps. Time stamps are formed when the transcriber highlights the time interval when one speaker or one signal is transmitted. The audio corresponding to the time stamps is then converted to text creating a complete strip of data. Hence, a time stamp represents the time intervals of a single transmission of one speaker at a time. They are mostly short during conversations and longer through broadcasts.

5.1.2 Data Generated and Analysis

A total of 27 hours of audio recordings are processed and converted from audio to text using word-by-word transcription. The transcription tool used is an open-source software called Subtitle edit (Niske, n.d.). Those 27 hours are broken down into 14 hours from the United States territorial waters, 5 from Canadian waters, 3 from the United Kingdom, 2 from Australia, and 2 from Europe.

The reference transcription outcome, i.e., characters identified and hence transcribed by the Marine Expert, is averaged at 66.40%. Hence out of the 27 hours processed, 66.40% was transcribed which corresponds to 18 hours of transcribed data. This value is estimated by the transcriber on a weekly basis and averaged at the end of the transcription phase of the experiment. An example is shown in Table 2 below.

Table 2. AU file example - Transcription quality reporting and averaging

Transcription %	AU File (3 hours)			
Date	Progress (Hours)	Total Hours	*Reported Quality (% transcribed)*	Hourly weighted quality
27-Sep	0.25	0.25	*75*	6.25
04-Oct	0.75	1	*65*	16.25
12-Oct	1	2	*65*	21.67
21-Oct	1	3	*60*	20.00
				64.17

Those 18 hours of transcribed data comprised 9 hours and 30 minutes from the United States audio file (67.14% of 14 hours), 3 hours and 30 minutes from the Canada file (71% of 5 hours), 2 hours from the United Kingdom file (68.33% of 3 hours), 2 hours from the Australia file (64.17% of 3 hours), and 1 hour from the Europe file (50% of 2 hours). Figures are illustrated in Table 3.

Table 3. Transcription Data Outcome Representation

Territory	Raw Data (Hours)	Successfully transcribed (%)	Weighted %	Transcribed Data (Hours)
US	14	67.14	34.81	9 hours 30 min
CA	5	71.00	13.15	3 hours 30 min
UK	3	68.33	7.59	2 hours
AU	3	64.17	7.13	2 hours
EU	2	50.00	3.70	1 hour
	27 Hours		**66.40%**	**18 Hours**

The volume of data generated can also be represented by the number of time stamps or alternatively lines. Total number of transcribed lines is 15300 with 6200 lines generated from the United States file, 3100 from the Australian file, 2500 from the Canadian file, 2200 from the United Kingdom file, and 1300 lines from the European file.

Table 4. Number of Time stamps by territory

Territory	Transcribed Data (Hours)	Number of Time stamps/Lines
US	9 hours 30 min	6200
CA	3 hours 30 min	3100
UK	2 hours	2500
AU	2 hours	2200
EU	1 hour	1300
	18 Hours	15300

The relatively low transcription outcome is due to the low quality of the audio files and the human factor. Low quality of recordings refers to the interference of radio signals that cause poor voice clarity, shakiness and distortion, and background noise. Not to mention the linguistic challenges established by radio operators such as low speech volume, heavy accents, and usage of foreign languages and local dialects by radio operators.

5.1.3 Broadcasts

Broadcasts referred to in this paper are long messages ‘more than 15 seconds’ transmitted by official bodies such as the Coast Guard, Vessel Traffic Services, Meteorological institutes, port authorities, MRCC, and so on. As shown in Table 5, broadcasts represent 4.3% of the transcribed data, which is equivalent to 46 minutes of speech.

Table 5. Broadcasts territorial comparative analysis

Territory	Transcribed Data (Hours)	Broadcasts %	Broadcasts
US	9 hours 30 min	7.06%	39.8 minutes
CA	3 hours 30 min	0.00%	0.0 minutes
UK	2 hours	2.45%	3.0 minutes
AU	2 hours	1.06%	1.2 minutes
EU	1 hour	3.69%	2.2 minutes
		4.3%	**46.3 minutes**

When it comes to the calculations, broadcasts in each file are counted in time stamps/lines. As transmissions are generally longer than conversational transmissions, an estimated factor of 1:3 is introduced. Consequently, the total number of broadcasts in each file, and later collectively, is multiplied by 3 to increase the reliability of the results. An illustration of the calculations is shown in Table 6 below.

Table 6. Broadcasts estimation calculations overview

Territory	Total Lines	Broadcast Lines	Estimation (x3)*	Broadcasts %	Broadcasts
US	6200	146	438	7.06%	**39.8 minutes**
CA	2500	0	0	0.00%	**0.0 minutes**
UK	2200	18	54	2.45%	**3.0 minutes**
AU	3100	11	33	1.06%	**1.2 minutes**
EU	1300	16	48	3.69%	**2.2 minutes**
	15300	**191**	**573**		**46.3 minutes**

5.1.4 Observations

It was observed that English-speaking countries had relatively higher transcription outcome percentages. They also averaged within a close range of 7% ‘min-max’, with Canada leading

at 71%, followed by the United Kingdom at 68.33%, the United States at 64.17%, and finally Australia with an outcome percentage of 64.17%.

On the other hand, the transcription outcome within European territorial waters was the lowest at 50%. That is 14% less than the lowest outcome percentage 'Australia' and 21% lower than the highest 'Canada'. This suggests that it is more difficult to detect and identify English speech in countries which use English as a second language.

Arguably, the total hours transcribed from Europe were 2 hours only. This shows that the data volume is low and hence is insufficient as a basis for assumption. Another observation explaining the low transcription outcome is, as Europe use English as a secondary operation language, numerous radio transmissions were not in English but instead in the local language of the respective territorial waters. Hence, such transmissions were disregarded and not transcribed.

On the other hand, when it comes to the Marine expert, a trend is noticed. A decrease in the performance of the Marine expert was attributed to heightened levels of exhaustion and pressure caused by the nature of the transcription job. Furthermore, a decrease in attentiveness and an increase in sedentary behaviour were noted after a certain period. Considering these observations, the transcription process was cautiously managed to regulate the workload and maintain the quality of the results.

5.2 Phase II: Conversion to SMCP

The second phase of building the specialized maritime library is the conversion of the transcribed text into Standard Maritime Communication Phrases 'SMCP'. This acts as the second set of data in the library. The goal is to familiarize the engine with the standard maritime protocol and domain terminology so that it can:

1. Analyse and correct speech anomalies

2. Translate from casual speech, and represent SMCP text on screen 'if needed'

An overview of how the SMCP set of data may be beneficial can be demonstrated in the following 2 scenarios. The engine theoretically transcribes an incoming transmission and generates X sentence. In the specialized library it, is given that the sentence X corresponds to the sentence Y in SMCP. Hence the engine displays both sentences X, and Y on the screen. This paves the way for a smart maritime speech recognition engine that can 'understand'

casual speech. Hence facilitating the job of the engine/feature end user by avoiding misunderstandings and so on.

The second scenario is the engine transcribing an incoming transmission as ‘Stop sah and rescue operations’. The engine can then analyse and generate the right message, which will be ‘Stop search and rescue operations’ in this example. Consequently, this can build up the foundation for the identification of casual radio speech and ‘translating’ it to Standard Marine Communication Phrases ‘SMCP’ for the speech recognition engine end user as mentioned. Last, this phase allows checking the transcripts in case of terminology anomalies and spelling mistakes made by the transcriber in phase 1.

5.2.1 SMCP Conversion process

To prepare the set of SMCP data, the transcribed text from the first phase is read and observed thoroughly. Next, an assessment is made to determine the stamps that can be converted to SMCP. After that, specific sections and sentences of the stamps are chosen to be worthy of conversion, before commencing with the conversion process. An illustration of the conversion process is shown in Table 7 below.

Table 7. Illustration of the SMCP Conversion process

Original text ⟶	SMCP
pilot ladder point five four meters	Pilot ladder is five decimal four meters above the water
can you shift the channel two two alpha	Switch to channel two two alpha
can you give me your *gps* positioning geographic location over	What is your GPS position
can you give me the description of your vessel	Report a brief description of your vessel
there are phone number that we can contact you	What is your phone number
requesting now your future inten intention	What are your intentions
copy that	Understood
i'm traffic in the moment here just standby	Please stand by. I am in a traffic situation
do you copy	Do you read me
do you guys hear the radio	Do you read me
requesting an onboard phone number	What is your on-board phone number
starboard side ladder please	Rig Pilot ladder on the starboard side
nine knots boarding speed	Pilot boarding speed is nine knots

5.2.2 Data Generated and Analysis

A total of 1674 SMCP lines are generated from the transcribed data lines with a weighted conversion ratio of 9.26%, which is equivalent to 1 hour and 40 minutes of recordings.

As illustrated in Table 8, those SMCP lines are broken down into 406 lines from the United States file with a conversion ratio of 6.55%. 612 lines from Australia file with a 19.74% conversion ratio. 163 lines from the Canada file with a 6.52% conversion ratio. 246 lines from the United Kingdom file with an 11.18% conversion ratio, and 247 lines from the Europe file with a 19% conversion ratio.

Table 8. Territorial Comparative Analysis of successful conversion to SMCP from transcribed data[1]

Territory	From Transcribed Data (Hours)	Successfully converted to SMCP (%)
US	9 hours 30 min	6.55
CA	3 hours 30 min	6.52
UK	2 hours	11.18
AU	2 hours	19.74
EU	1 hour	19.00
	18 Hours	

To obtain the overall SMCP conversion rate, each file's SMCP % percentage is used against the respective transcribed hours to deduce the hypothetical length of each SMCP file in minutes as shown in Table 9. For instance, 6.55% of the 9.50 transcribed hours of the US file equals to 37.33 minutes of recordings converted to SMCP. All the values are added and the total SMCP time in minutes is found to be 100 minutes which is equivalent to 1 hour 40 minutes. These 100 minutes are then divided by the initial total transcribed hours (18) to obtain the overall SMCP conversion rate of **9.26%.**

[1] *Broadcast messages transmitted by official bodies such as the coast guard, Vessel Traffic Services, Meteorological institutes, etc. are not included*

Table 9. Deducing overall SMCP conversion rate: Data converted to SMCP ‘in minutes’

File	Total lines	SMCP lines	SMCP %	Total Hours	Converted (t min)
US	6200	406	6.55%	9.50	37.33
AU	3100	612	19.74%	3.50	13.69
CA	2500	163	6.52%	2.00	13.42
UK	2200	246	11.18%	2.00	23.69
EU	1300	247	19%	1.00	11.40
	15300	**1674**		**18**	**99.53**

For an overview, Figure 3 below gives a summary of how the volume of data shrunk as the process of transcription and conversion progressed. The process initiated with 27 hours of audio recordings, where 18 are transcribed. Out of these 18 hours, only 1.7 hours are converted to SMCP.

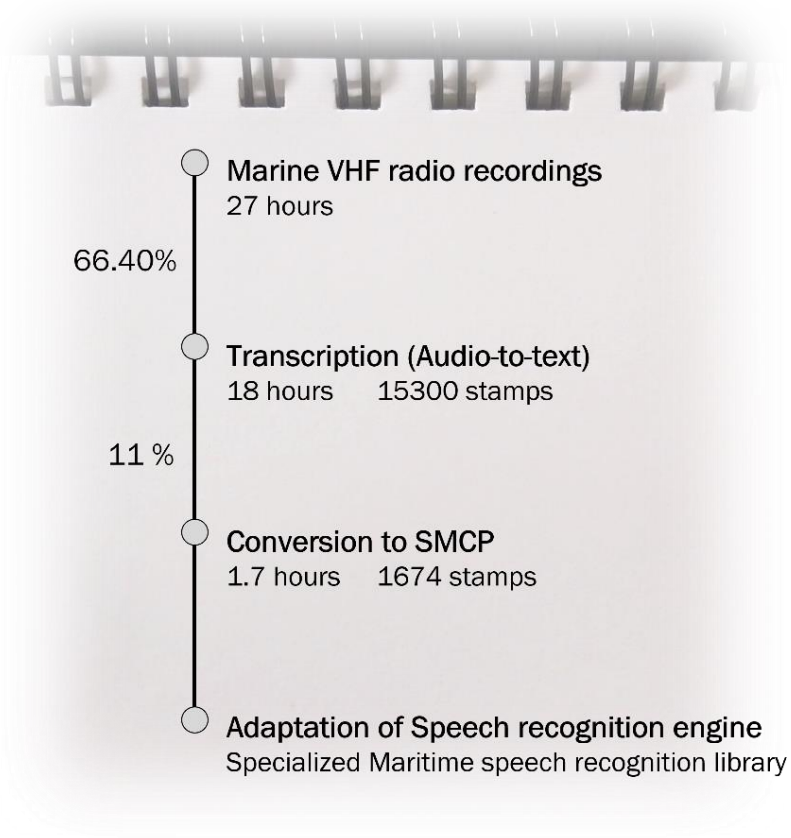

Figure 3. Data shrinkage: Process of the library building

5.2.3 Challenges, comments, and observations

The process of analysing the transcribed data, and consequently deciding on the sections to be converted to SMCP has its challenges. For instance, the speech can accommodate multiple meanings depending on the context and background information which is vital as stated by Lazaraton (according to Dževerdanović-Pejović, 2013, IJTTE, p. 378), yet inaccessible in many cases. Furthermore, the degree of compliance with IMO SMCP protocol is questionable as the personnel handling the conversion process are students.

An observation regarding the SMCP is that it emphasizes on distress and urgency protocols. This is also noted in (Ahmmed, 2017) as it is stated that the SMCP phrases "are mostly related to berthing and responses in emergency situation" (Ahmmed, 2017, BMJ Volume 1 Issue 1, p. 29). In the current approach, only the operational protocols were deemed viable as the recordings rarely comprised of any distress or urgency scenarios.

5.3 Phase III: Fine-tuning

This chapter illustrates how the newly built specialized maritime data library is utilized in adapting or fine-tuning the general-purpose speech recognition model by a third party, Lingsoft Oy (Lingsoft Oy, n.d.) throughout the end of 2022 and January 2023. The information presented in this chapter is mainly acquired from the interview with Michael Stormbom, the Chief Technical Officer at Lingsoft Oy as mentioned in the Appendix, and the person in charge of the speech recognition engine adaptation.

The initial challenge was to build the specialized maritime speech recognition data library. This process required annotating the maritime-specific data, which included terminologies unique to the industry, such as ship names and the Standard Marine Communication Phrases 'SMCP'. Once the library is built, as shown in the previous chapters, the fine-tuning process is initiated and involved fine-tuning the model by adding the new data to generate a maritime-specific model.

The library is deemed "pretty good overall" (Interview with Stormbom, 6 March 2023). Although the annotations could have been improved during the first stage of the transcription process. Basically, when the radio VHF recordings were converted from voice-to-text manually by a Marine expert, some areas labelled as 'Unknown' by the Marine expert could have been transcribed to the correct text as the respective audio sections were not challenging. (Interview with Stormbom, 6 March 2023).

According to my observations, this drop in the performance of the Marine expert was due to the increasing levels of fatigue and stress due to the nature of the work. After a certain amount of time declining focus and sedentary behaviour were observed. Therefore, the workload was carefully controlled during the transcription process from that point onwards.

Moving forward, one of the key challenges in the adaptation process highlighted in the interview is the difficulty in finding the right model to adapt. The interviewee explained that they tried with one general-purpose English AI model initially, but the results were not satisfactory. This highlights the importance of selecting the right base model to adapt and the need for experimentation and trial-and-error. However, when a suitable model was found, the results were significantly better.

Another aspect is the challenge of acquiring the data and getting it into the right shape, as AI models rely on data to function, which accordingly affects their behaviour and output. Hence the acquisition, analysis and refinement of the maritime fine-tuning data present a challenge and is an area to be studied further (Interview with Nakilcioglu 5 April 2023) (Interview with Stormbom, 6 March 2023). Furthermore, the poor quality of the radio VHF audio recordings, which resulted from the overall quality of the radio signal, also contributed to the challenges of Lingsoft at this stage.

In conclusion, given the base model is suitable and the data is of satisfactory quality and volume, the process of fine-tuning or adaptation is deemed straightforward and of “moderate” difficulty (Interview with Stormbom, 6 March 2023). It is to be noted that only the first part of the specialized data library is utilized for the fine-tuning process, that is, the manual transcription with the corresponding time stamps forming data labels (Interview with Nakilcioglu 5 April 2023).

The second part, which consists of the converted transcriptions to SMCP, is not utilized in this experiment. The sentences of converted SMCP text cannot contribute to the fine-tuning of the speech recognition model. However, it can be utilized in the post-processing stage later or Natural language processing, if applicable. (Interview with Nakilcioglu 5 April 2023).

6 Results and interpretation of the results

The process of testing the engines, along with the results generation and empirical analysis are shown in this chapter. The results are analyzed using the three factors mentioned in

Chapter 4 of this paper, transcription outcome, transcription accuracy and WER. A walkthrough of the calculations and possible error corrections for all factors are demonstrated in this chapter as well.

6.1 Testing

Testing procedures involve running of both the Specialized maritime speech recognition engine, and the standard speech recognition engine with sample data. Results or transcripts generated by the Marine expert are used as a control for the experiment to allow for a fair comparison.

As mentioned earlier, the sample data is a total of 20 minutes of maritime VHF radio recordings, broken down into about 4 minutes of pilot boarding operations communications, around 11 minutes of weather forecasts, warnings, coast guard messages, and urgency 'sécurité transmissions, and finally 2 and 3 minutes of random sets of audio data from the United States, and Australia respectively. The length of each recording is shown in Table 10 below.

Table 10. Sample data specifications

Sample file	Total length
Pilot	3 minutes 36 seconds
Broadcasts	11 minutes 16 seconds
Random - US	2 minutes
Random - AU	3 minutes 12 seconds
	20 minutes 4 seconds

6.2 Transcription Outcome

The transcription outcome of an engine is presented as a percentage of the Marine expert outcome. A marine expert is deemed to identify and transcribe 100% of the audio recordings. This assumption is made to facilitate the comparative analysis of the results. The actual transcription outcome of the Marine expert is 66.40% as presented earlier in Table 2 under Chapter 4.1.2 Data generated and analysis.

The transcription outcome of respective engines is calculated based on the total number of words generated after running the test samples. The number of words is referred to as Total engine character count. Table 11 shows the character count generated by the Marine expert, the specialized maritime speech recognition engine, and the standard speech recognition engine.

Table 11. Total engine character count: Number of words generated after testing

Samples	Marine Expert	Maritime SR engine	Standard engine
Pilot	331	300	285
Broadcasts	1523	1049	759
Random AU	210	137	154
Random US	461	461	460
	2525	**1947**	**1658**

The Marine expert generated 2525 characters, which corresponds to 100% of the transcription outcome as stated earlier. Respectively, the specialized maritime speech recognition engine generated 1947 characters, which is higher than the 1658 characters generated by the standard speech recognition engine.

Using a basic arithmetic approach and through cross multiplication, as the 2525 characters generated by Marine expert correspond to 66.40%, and the Maritime SR engine generated 1947 characters, a transcription outcome of 51.20% can be deduced. The 66.40% is the Marine expert's transcription outcome from Table 2.

Table 12. Deduced Transcription rate for marine specialized speech recognition engine

Transcription Outcome	Character Count
66.40%	2525
51.20%	1947

The same approach applies to the standard speech recognition engine. Hence, inducing a lower transcription outcome of 43.60%. The trend has shown that the transcription outcome

of the specialized maritime speech recognition engine is higher than the standard speech recognition engine. The adapted engine can identify and transcribe more characters compared to the standard one as graphically represented in Figure 4.

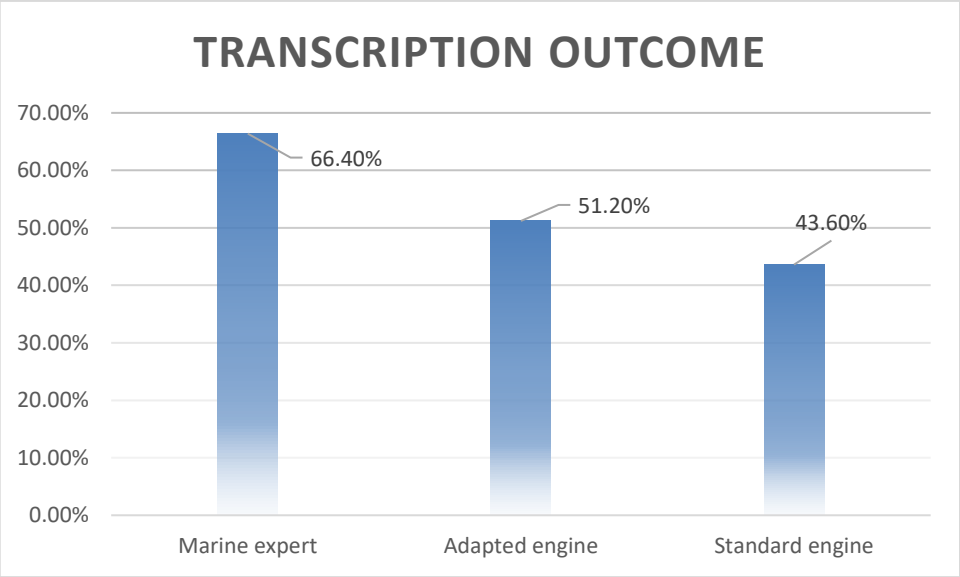

Figure 4. Transcription Outcome of Engines

6.3 Transcription Accuracy

As mentioned earlier in Chapter 4, Transcription accuracy indicates how correct the characters generated by the speech recognition engine are. This chapter highlights the significance of accuracy in maritime communication and the possible problems in case of misinterpretation. Furthermore, it dives into the process of analysing the engines' output to determine their respective accuracies before concluding on the findings.

6.3.1 Accuracy Relevancy

Accuracy of the transcripts is an important factor that directly affects the quality, and consequently the efficiency of the communication. This directly contributes to the safety of the respective working operation regardless of its nature. In the maritime domain, a VHF radio transmission is usually concise and passes on a specific piece of information using maritime terminology and the Standard Maritime Communication Phrases 'SMCP'.

The language models of general-purpose speech recognition engines do not possess the maritime terminology and the domain speech patterns, let alone the acoustic challenges. Hence, they are unable to detect and transcribe the VHF maritime transmissions accurately, especially the industry-specific terms. This stands as an operational limitation as the core of

the message might be one or two words only. The following paragraphs illustrate an example of a VHF radio transmission that is proved unworthy because of the standard engine not being able to detect and correctly transcribe the 'core' piece of information of the message.

The following are actual transcripts of recordings from the US audio file encompassing a series of pilot boarding instructions transmitted by a Pilotage services radio operator to an approaching vessel. Pilot boarding instructions inform the vessel on five pieces of operational information contributing to the safe and efficient boarding of the pilot.

1. Pilot boarding time
2. Pilot boarding speed
3. Pilot ladder side
4. Pilot ladder height above the water
5. Pilot boat working radio VHF channels

The transcripts are generated by the Marine Expert and the Standard Automatic Speech Recognition engine respectively. The aim of this illustration is to compare the accuracy of the two approaches, provided the outcome percentage is almost the same.

1. Marine Expert

'Sir just confirming pilot onboard this morning zero zero six zero zero local time starboard side ladder please two meters off the water nine knots boarding speed of the pilot station sir. Then pilot boat will be standing by on one three and one zero one zero working channel all the time'

2. Standard Automatic Speech Recognition engine

'Sir. Uh, just confirming pilot on board this morning. 0 6 0 0. Local time. Uh, stop side ladder please. Two meters off the water. Nine knots boarding speed at the pilot station, sir. Then, uh, pilot will be standing by on 13 working channel all the time'

'stop' - The anomaly in the Standard Automatic Speech Recognition engine transcript, though one word, deems the instruction invalid and is hence disregarded. As per the Standard Maritime Communication Phrases 'SMCP', the pilot ladder can be either on the port side or

the starboard side and therefore the outcome 'stop' is an invalid entity. As a result, this information is discarded and will have to be requested again by the approaching vessel.

'one zero one zero' – The Standard Automatic Speech Recognition engine is unable to detect the Pilot boat second working VHF radio channel.

Table 13 below briefly illustrates the impact of the standard engine transcript's accuracies towards the efficiency of communication and consequently the safety of working operations.

Table 13. Qualitative comparison between ASR engine and marine expert's transcription outcomes

Marine expert	ASR engine	Comments
… starboard side ladder …	… stop side ladder …	Ship's side is either port or starboard
… pilot will be standing by on one three and one zero one zero working channel all the time …	… pilot will be standing by on 13 working channel all the time …	The second working channel was not detected
… port side …	… court side …	Ship's side is either port or starboard

6.3.2 Data Generated and Analysis

The transcription accuracy will also be presented as a percentage of the Marine expert's accuracy. A marine expert is deemed to correctly transcribe 95% of the characters in an audio recording. The accuracy of the engines will be interpolated to the 95% of the Marine expert at the end. This assumption is made to facilitate the comparative analysis of the results.

As per Table 14, the *initial text similarity percentage* between the Marine expert transcripts and the specialized maritime speech recognition engine for the pilot file transcript is 86.00%, 77% for the broadcasts file, and 71% for the Random AU file. On the other hand, the standard speech recognition engine transcripts had an initial similarity percentage of 70%, 50% and 63% for the pilot, broadcasts, and the Random AU transcripts respectively.

Table 14. Transcription accuracy - Initial text similarity percentage

	Initial similarity percentage	
Samples	Standard engine	Adapted engine
Pilot	70.00%	86.00%
Broadcasts	50.00%	77.00%
Random AU	63.00%	71.00%

Please note that the transcripts of the 4th sample file 'Random US' are not analysed towards the calculation of the transcription accuracy due to technical difficulties. The text analysis and similarity tool is unable to provide a similarity percentage for the maritime specialized speech recognition engine Random US file transcript.

6.3.3 Number format error correction

Next, the number format error is noticed. The standard speech recognition engine transcribed the numbers as digits '1, 56, 24' instead of in the textual form. This error had to be taken into consideration and corrected to have a fair comparison. This is done in three steps.

As a reminder, the first step is to extract all the numbers from the transcripts, then these digits are counted. Finally, the number of *Accurate digits* is obtained by multiplying the *Number of digits* from the previous step with the respective sample's *initial similarity percentage* from Table 14. The findings are illustrated below in Table 15 for the standard engine.

Table 15. Standard engine - Accurate digits by sample file

Samples	Initial similarity %	Number of digits	Accurate digits
Pilot	70.00%	41	28
Broadcasts	50.00%	170	85
Random AU	63.00%	11	7

Consequently, the same is done with the rest of the text to get the Accurate words, in correspondence with Accurate digits in the previous step. The Total engine character count generated by the engines from Table 11 is multiplied by the *initial similarity %* from Table 14. This deduced number of Accurate words is added to the *Accurate digits* to get the *Total matching characters*. Table 16 below showcases the process.

Table 16. Standard engine - Total matching characters: Number format error correction

Samples	Initial similarity %	Accurate characters	Accurate digits	Total matching characters
Pilot	70.00%	200	28	228
Broadcasts	50.00%	380	85	465
Random AU	63.00%	97	7	104

Finally, a simple comparison between the *Total matching characters* and the *Total engine character count* from Table 11 gives the *Corrected similarity percentage* for a single engine as illustrated in Table 17 for the standard engine. This corrected similarity percentage is the transcription accuracy of the engine by sample.

Table 17. Standard engine - Corrected similarity percentage by sample

Samples	Total engine character count	Total matching characters	Corrected similarity %
Pilot	285	228	80.07%
Broadcasts	759	465	61.20%
Random AU	154	104	67.50%

Hence the updated transcription accuracy figures would be as shown in Table 18. It can be observed this the correction varies within the different samples. This is coherent as the *initial similarity %* and the *number of digits* extracted vary between the samples.

Table 18. Transcription accuracy – Corrected similarity percentage

	Corrected similarity percentage	
Samples	Standard engine	Adapted engine
Pilot	80.07%	86.00%
Broadcasts	61.20%	77.00%
Random AU	67.50%	71.00%

6.3.4 Interpolation and Conclusion

As stated earlier at the beginning of this chapter, the transcription accuracy is presented as a percentage of the Marine expert's accuracy which is assumed to be 95%. Hence, the transcription accuracies deduced and corrected earlier are to be interpolated to the 95% Marine expert accuracy. Based on the interpolated accuracies and the total engine character count, the weighted accuracies are obtained. Finally, the sum of the weighted accuracies produces the final transcription accuracy of the engine.

The transcription accuracy of the standard speech recognition engine is found to be **63.17%** while the specialized maritime speech recognition engine demonstrated an accuracy of **74.40%**. Table 19 shows the final interpolated accuracy of the Standard engine, with respect to the sample files, along with the weighted accuracies and the final transcription accuracy.

Table 19. Standard engine – Interpolation and final transcription accuracy

Samples	Corrected similarity percentage	Interpolated Accuracy %	Total engine character count	Weighted Accuracy %
Pilot	80.07%	76.07%	285	0.18
Broadcasts	61.20%	58.14%	759	0.37
Random AU	67.50%	64.13%	154	0.08
			1198	63.17%

Subsequently, Table 20 shows the final interpolated accuracy of the adapted engine, with respect to the sample files, along with the weighted accuracies and the final transcription accuracy. This serves as a representation of the empirical data and allows for the comparison of transcription accuracy values.

Table 20. Adapted engine – Interpolation and final transcription accuracy

Samples	Corrected similarity percentage	Interpolated Accuracy %	Total engine character count	Weighted Accuracy %
Pilot	86.00%	81.70%	300	0.16
Broadcasts	77.00%	73.15%	1049	0.52
Random AU	71.00%	67.45%	137	0.06
			1486	**74.35%**

The test results show that the adapted engine transcripts are more accurate than the standard engine transcripts. This suggests that the maritime speech recognition library built and used to adapt the speech recognition engine is worthy and generated more accurate characters or words compared to the non-adapted engine.

Arguably, the sample used to test the above hypothesis is small. Hence, the reliability of this claim is low. More tests are to be carried out in the future using maritime audio recordings of varying scenarios and difficulty to support the proposed claim.

To conclude, and as shown below in Figure 5, the standard speech recognition engine generated an outcome of 43.60% (Figure 4) with an accuracy of 63.20% (Table 19). On the other hand, the marine-adapted speech recognition engine shows an improved performance with an outcome of 51.20% (Figure 4) and an accuracy of 74.40% (Table 20) while the Marine expert had an outcome and accuracy of 66.40% (Figure 4) and 95.00% respectively.

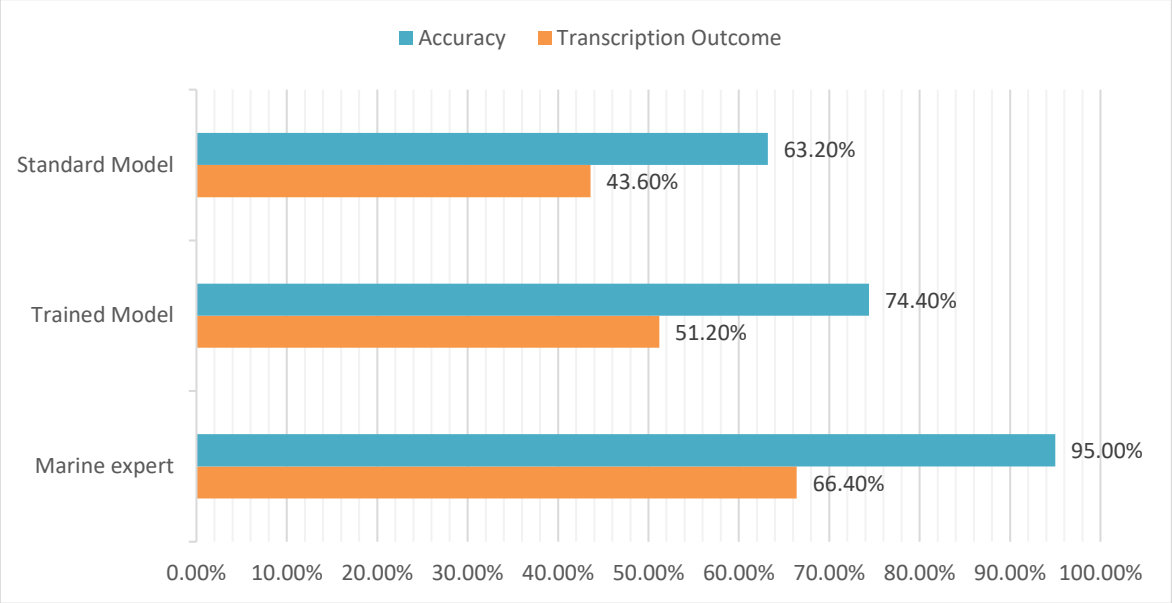

Figure 5. Conclusive comparison of the transcription outcome and accuracy

6.4 Word Error Rate (WER)

This chapter illustrates the process of obtaining the WER using the two approaches explained in Chapter 4. The first one uses the online WER calculator Amberscript, and the other uses the Python package JiWER.

Amberscript process is a straightforward one where the reference transcript 'Marine expert' is added to the *reference text box*, the respective engine transcript is added to the *automatic transcription text box* and the tool generates the WER. This comparison against the respective Marine expert transcript is done for every sample and for both engines, with results recorded. Table 21 below shows the WER results of the Amberscript tool.

Table 21: Amberscript WER figures

Samples	Standard engine	Adapted engine
Pilot	40.70%	35.20%
Broadcasts	66.70%	60.60%
Random AU	59.00%	57.10%

On the other hand, the Python approach is a bit troubling. After Poetry is installed and added to path successfully using Windows Powershell and the instructions on (Python Poetry, n.d.), the Python package JiWER is installed by following the steps on (Python Package Index PyPI, 2023). Then the code is compiled and the WER function is imported. Next, the *reference* transcript text is added, along with the respective *hypothesis* transcript text. Finally, the code is executed. The code used to do the calculation is shown below.

Code 1. JiWER: Computing WER

```
from jiwer import wer
reference = "Sample X Marine expert transcript"
hypothesis = "Sample X Engine transcript"
error = wer(reference, hypothesis)
print("Word Error Rate: ", error)
```

The above module is allowed to run for every sample, and for both engines, against the respective Marine expert transcripts. The module results are shown in Code 2 below. It can be observed that one test 'Adapted Random AU' generated a WER of 1.0 or 100%. This is due to human error. The error is noticed and acted upon. New WER deduced is 0.57 or 57.62%.

Code 2. JiWER: WER Module code

```
RESTART: C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Adapted Pilot.py
Word Error Rate:  0.38485804416403785
>>>RESTART:C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Adapted Broadcasts.py
Word Error Rate:  0.6293845135671741
>>>RESTART:C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Adapted Random AU.py
Word Error Rate:  1.0
>>>RESTART:C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Adapted Random AU.py
Word Error Rate:  0.5761904761904761
>>>RESTART: C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Standard Pilot.py
Word Error Rate:  0.5141955835962145
>>>RESTART: C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Standard Broadcasts.py
Word Error Rate:  0.7696889477167439
>>>RESTART: C:/Users/Ahmed/AppData/Local/Programs/Python/Python37/Scripts/Standard Random AU.py
Word Error Rate:  0.6904761904761905
```

The JiWER module results can be better presented in Table 22**.** Moreover, and like the Amberscript WER figures in Table 21, there is a large variation in the JiWER module results. For instance, highlighting the marine-adapted engine, the lowest WER is 38.50% for the pilot sample, and the highest is 62.94% for the broadcast sample. This indicates a lower reliability of the results.

Table 22. JiWER: WER figures

Samples	Standard engine	Adapted engine
Pilot	51.42%	38.50%
Broadcasts	76.97%	62.94%
Random AU	69.05%	57.62%

Finally, although the variation in the figures is abundant, and some of the tests generated exceptionally high WER values, and although two different WER tools are used, a trend is evident. The adapted engine always features a lower WER. Please be reminded that the lower the WER, the higher the accuracy of the engine, and consequently the better the performance of the system. This comparison is better shown in Table 23. It summarizes the findings of this WER chapter.

Table 23. Amberscript and JiWER: WER figures representation and comparison

	Amberscript		JiWER	
Samples	Standard engine	Adapted engine	Standard engine	Adapted engine
Pilot	40.70%	35.20%	51.42%	38.50%
Broadcasts	66.70%	60.60%	76.97%	62.94%
Random AU	59.00%	57.10%	69.05%	57.62%

7 Critical review and discussion

The aim of this paper is to assess the performance of a speech recognition engine after adapting it with a specialized maritime speech recognition data library, so as to assess the worthiness of introducing a specialized maritime speech recognition library.

In conclusion, the results of this thesis demonstrate the potential of specialized maritime speech recognition data libraries in improving the performance of general-purpose speech recognition engines in the maritime industry. The comparative analysis showed that the

marine-trained speech recognition engine outperformed the standard one in terms of transcription outcome, accuracy and WER, indicating that the library has a positive impact on the performance of the engine and its recognition capabilities. Hence the need to develop a standardized maritime data library is suggested.

The business potential of introducing speech recognition technology in the maritime domain is high, with a big market ready to be explored (Interview with Nakilcioglu 5 April 2023). Not to mention the expected improvement in operational safety aspects (Porathe, Eklund, & Göransson, 2021; Reimann & John, 2020). The results of this research can pave the way for the development of VHF radio communication decision-support and decision-making tools for shore-based remote operation centres.

The improved outcome and accuracy of the transcription offered by the marine-trained speech recognition engine can enhance the efficiency and safety of maritime operations, reduce the risk of human error, and improve the response time in emergency situations (Reimann & John, 2020). Further research is needed to explore the full potential of this method of adaptation, the volume and quality of the maritime data library, and the possible applications in the industry.

A proposal to improve the speech recognition engine results is the introduction of marine domain rules. These rules can act as a post-processing refinement measure for the engine output, like natural language processing or NLP. For instance, the engine transcribes an incoming transmission as ‘XXXX XXX Heading 445 degrees’ which is invalid and has no meaning. Instead, the introduction of a domain rule stating that ‘Heading between 000 and 359 degrees’ can increase the reliability of the engine transcripts

On the other hand, one challenge that this adaptation process has is the acquisition and processing of high-quality and diverse speech data for building the specialized data library (Interview with Nakilcioglu 5 April 2023; Interview with Stormbom, 6 March 2023). According to (Interview with Stormbom, 6 March 2023), approximately two or three times the current library is needed to achieve satisfactory results.

However, there are limitations and shortcomings that need to be addressed regarding the methodology of this research. One of the key limitations is the relatively small volume of the specialized maritime data library built and utilized for this experiment. This is due to the constraints in human factors, resources, and time. Also, it did not include a lot of scenarios from the maritime domain such as a Man overboard or a grounding situation for instance.

Moreover, it was built entirely by students and hence its quality is also questionable, especially the manual transcription part. The students are entry-level transcribers and acquire a relatively small experience of on-board experience and exposure to maritime VHF radio communication. Not to mention the challenging nature of the work which leads to fatigue, stress and consequently an expected drop in the quality of the transcripts with time.

This, in my opinion, contributes to the large WER figures, and the big variance observed in this experiment. Hence suggesting that the shortcomings and mistakes might have been in the reference text and not in the engine results. Unfortunately, no quality assessment was performed on the generated data by a professional marine transcriber. Instead, (Interview with Stormbom, 6 March 2023) highlighted the possibility of improvement in the annotations during the initial manual transcription process by the students.

Another factor that may have contributed to the high WER figures and the notable variance is the possible differences in punctuation marks such as hyphens, commas and so on between the comparative transcripts. This is possible as different engines generate transcripts in different formats. The idea sparked after observing that the JiWER function is sensitive to capitalization, unlike Amberscript. However, this assumption could not be confirmed and has no clear evidence.

Furthermore, the sample data was only 20 minutes long which is a relatively small test sample. The level of difficulty of the samples, although varied from one recording to another, was mostly easy to moderate and not very challenging as per my personal observation. These arguments address the limitations of the research methodology that might affect the reliability of the results.

In general, the performance of the speech recognition engine may vary depending on factors such as background noise, accent, and speech rate. Therefore, further research and development are needed to improve the robustness and adaptability of the engine for different communication scenarios in the maritime industry.

Commercially, there are still challenges that need to be addressed before these tools can be widely adopted such as the integration of these tools with existing communication systems and compatibility with international regulations and standards. Not to mention the concerns of stakeholders related to data privacy, cybersecurity, and human factors.

Overall, this thesis provides a foundation for the development of VHF radio communication decision-support and decision-making tools for shore-based Remote Operation Centres.

While challenges in data library creation and models' fine-tuning exist, the use of specialized data libraries and fine-tuning the existing models offer a promising avenue for future research and development.

8 Reference list

Ahmmed, R. (2017). The Difficulties of Maritime Communication and the Roles of English Teachers. *Bangladesh Maritime Journal (BMJ), 1*(1), 22-34.

Allianz Global Corporate & Specialty. (2012). *Safety and Shipping 1912-2012.* Retrieved January 27, 2023, from https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/migration/media/press/document/other/agcs_safety_shipping_1912-2012.pdf

Amberscript. (n.d.). *Word Error Rate Tool: Measure automatic speech recognition accuracy objectively*. Retrieved April 16, 2023, from Amberscript Web site: https://www.amberscript.com/en/wer-tool/

Anantaram, C., & Kopparapu, S. K. (2017, October). Adapting general-purpose speech recognition engine output for domain-specific natural language question answering. doi:10.48550/arXiv.1710.06923

Avasthi, A. (2021, December 3). Using NLP for Automatic Speech Recognition. datasaur.ai Blog post. Retrieved March 23, 2023, from https://datasaur.ai/blog-posts/nlp-speech-recognition

Barreto, S. (n.d.). *What is Fine-tuning in Neural Networks?* Retrieved April 19, 2022, from Baeldung Web site: https://www.baeldung.com/cs/fine-tuning-nn

broswerling. (n.d.). *Extract All Numbers From Text*. Retrieved March 16, 2023, from broswerling Web site: https://www.browserling.com/tools/extract-numbers

Chrysanthi, P. (2019, July 30). *Safety challenges for the human factor in the maritime industry*. Retrieved February 17, 2023, from SAFETY4SEA: https://safety4sea.com/cm-safety-challenges-for-the-human-factor-in-the-maritime-industry/

CommonVoice Mozilla. (n.d.). *About CommonVoice*. Retrieved April 12, 2023, from CommonVoice Web site: https://commonvoice.mozilla.org/en/about

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). Unsupervised Cross-Lingual Representation Learning for Speech Recognition. *Proceedings of Interspeech 2021.* doi:10.21437/Interspeech.2021-329

Descript. (n.d.). *Descript Homepage*. Retrieved February 15, 2023, from Descript Web site: https://www.descript.com/

Dževerdanović-Pejović, M. (2013). DISCOURSE OF VHF COMMUNICATION AT SEA AND THE INTERCULTURAL ASPECT. *International Journal for Traffic and Transport Engineering, 3*(4), 377-396. doi:10.7708/ijtte.2013.3(4).03

ELNAV Advanced Safety Systems. (n.d.). *Helm Order Monitor*. Retrieved February 6, 2023, from ELNAV Corporation Web site: https://elnav.ai/helm-order-monitor/

ELNAV. (n.d.). *Homepage: ELNAV*. Retrieved February 6, 2023, from ELNAV Corporation Web site: https://elnav.ai/

ExtendOffice. (n.d.). *How to merge or combine multiple lines into a single paragraph in Word document?* Retrieved April 4, 2023, from ExtendOffice Web site: https://www.extendoffice.com/documents/word/5413-word-merge-multiple-lines.html

Fox, D. (2021, September 9). *Is Word Error Rate Useful?* Retrieved April 13, 2023, from AssemblyAI Web site: https://www.assemblyai.com/blog/word-error-rate/

Fraunhofer CML. (n.d.). *Automated Transcription of Maritime VHF Radio Communication for Search and Rescue (SAR) Mission Coordination*. Retrieved February 6, 2023, from Fraunhofer CML: https://www.cml.fraunhofer.de/en/research-projects/ARTUS.html

Fraunhofer IDMT. (2022, November 23). *Automatic voice monitoring for the ship's bridge*. Retrieved February 6, 2023, from Fraunhofer IDMT Corporate Web site: https://www.idmt.fraunhofer.de/en/Press_and_Media/press_releases/2022/automatic-voice-monitoring-at-sea.html

Guillaume, S., Wisniewski, G., Macaire, C., Jacques, G., Michaud, A., Michaud, A., . . . Fily, M. (2022). %T Fine-tuning pre-trained models for Automatic Speech Recognition, experiments on a fieldwork corpus of Japhug (Trans-Himalayan family). *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 170-178). Dublin, Ireland: Association for Computational Linguistics. doi:10.18653/v1/2022

Happyscribe. (n.d.). *Happyscribe: Free Transcription Software*. Retrieved January 5, 2023, from Happyscribe Corporation Web site: https://www.happyscribe.com/free-transcription-software

IBM. (n.d.). *What is speech recognition?* Retrieved April 6, 2023, from IBM Corporation Web site: https://www.ibm.com/topics/speech-recognition

IMO. (2001, November). *IMO SMCP Resolution A.918(22).* Retrieved February 4, 2023, from International Maritime Organization website: https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/A.918(22).pdf

International Chamber of Shipping. (2021). *Seafarer Workforce Report.* International Chamber of Shipping. Retrieved February 2, 2023, from https://www.ics-shipping.org/publication/seafarer-workforce-report-2021-edition/

International Chamber of Shipping. (2021). *Shipping and World Trade: Global Supply and Demand for Seafarers*. Retrieved February 2, 2023, from International Chamber of Shipping Web site: https://www.ics-shipping.org/shipping-fact/shipping-and-world-trade-global-supply-and-demand-for-seafarers/

International Maritime Organization (IMO). (2021, May 25). IMO's Maritime Safety Committee finalizes its analysis of ship safety treaties, to assess next steps for regulating Maritime Autonomous Surface Ships (MASS). Retrieved February 4, 2023, from https://www.imo.org/en/MediaCentre/PressBriefings/pages/MASSRSE2021.aspx

International Maritime Organization (IMO). (n.d.). *IMO Standard Marine Communication Phrases*. Retrieved February 4, 2023, from International Maritime Organization (IMO) Web site: https://www.imo.org/en/ourwork/safety/pages/standardmarinecommunicationphrases.aspx

International Maritime Organization IMO. (2010). *International Convention on Standards of Training, Certification and Watchkeeping for Seafarers STCW Manila 2010.* Retrieved February 4, 2023

Kataria, A. (2011). Maritime English and the VTS. *International Maritime English Conference IMEC 23*, (pp. 25-33). Constanta, Romania.

Kurimo, M. (1997, July 11). Recognition error rate. Retrieved February 3, 2023, from https://users.ics.aalto.fi/mikkok/thesis/book/node37.html#:~:text=The%20error%20rate%20consists%20of,all%20tested%20speakers%20are%20averaged

Lingsoft Oy. (n.d.). *Home page: Lingsoft Oy*. Retrieved November 13, 2022, from Lingsoft Oy Corporation Web site: https://www.lingsoft.fi/en

Niske. (n.d.). *Subtitle edit*. Retrieved October 12, 2022, from Niske dk: https://www.nikse.dk/subtitleedit

Noble, A. (2007). The IMO SMCP 15 years on: current perceptions and realistic recommendations. *IMLA-International Maritime English Conference (IMEC27)*, (pp. 127-145).

oTranscribe. (n.d.). *oTranscribe Main page*. Retrieved February 4, 2023, from oTranscribe Web site: https://otranscribe.com/

Otter.ai. (n.d.). *Otter.ai - Voice Meeting Notes & Real-time Transcription*. Retrieved February 4, 2023, from Otter.ai Corporation Web site: https://otter.ai/

Partanen, N., Hämäläinen, M., & Klooster, T. (2020). Speech Recognition for Endangered and Extinct Samoyedic languages. *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation* (pp. 523-533). Hanoi, Vietnam: The Association for Computational Linguistics.

Pekichev, P. (2021, September 14). Post-processing in automatic speech recognition systems. Retrieved December 5, 2022, from https://blog.webex.com/engineering/post-processing-in-automatic-speech-recognition-systems/

Perez, D., & Manuel, J. (2003). IMO Standard Marine Communication Phrases and teaching their use in VTS-context. *Bulletin de l'AISM = IALA bulletin*, 20-33.

Picovoice.ai. (2022, October 4). NLP Applications in Voice Recognition. Retrieved April 9, 2023, from https://picovoice.ai/blog/voice-recognition-NLP/#:~:text=NLP%20and%20Voice%20Recognition%20are,cannot%20directly%20process%20audio%20inputs.

Porathe, T., Eklund, P., & Göransson, H. (2021). Voice and Text Messaging in Ship Communication. *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE (2021).* doi:10.54941/ahfe100620

Pritchard, B. (2003). Maritime English Syllabus for the Modern Seafarer: Safety-related or Comprehensive Courses? *WMU Journal of Maritime Affairs, 2*(2), 149–166. doi:10.1007/BF03195041

Prud'hommeaux, E., Jimerson, R., Hatcher, R., & Michelson, K. (2021). Automatic Speech Recognition for Supporting Endangered Language Documentation. *Language Documentation & Conservation*, *15*, pp. 491-513.

Python Package Index PyPI. (2023, March 28). *jiwer 3.0.1*. Retrieved April 14, 2023, from Python Package Index PyPI Web site: https://pypi.org/project/jiwer/

Python Poetry. (n.d.). *Introduction*. Retrieved April 14, 2023, from Poetry: PYTHON PACKAGING AND DEPENDENCY MANAGEMENT MADE EASY: https://python-poetry.org/docs/

Python Software Foundation. (n.d.). *Python Home page*. Retrieved April 14, 2023, from Python Software Foundation Web site: https://www.python.org/

Reimann, M., & John, O. (2020). Increasing Quality of Maritime Communication through Intelligent Speech Recognition and Radio Direction Finding. *European Navigation Conference (ENC) 2020.* Dresden, Germany. doi:10.23919/ENC48637.2020.9317332

Ringger, E. K., & Allen, J. F. (1996). Error Correction Via A Post-Processor For Continuous Speech Recognition. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings.* Atlanta, GA, USA. doi:10.1109/ICASSP.1996.541124

Schriever, U. (2009). Acceptance of, oppposition to and competency levels in Maritime English as seen by seafarers. *Proceedings of International Maritime English Conference (IMEC21)*, (pp. 151-164). Szczecin, Poland.

Temi. (n.d.). *Homepage: Temi*. Retrieved January 7, 2023, from Temi Corporation website: https://www.temi.com/

the FTW Transcriber. (n.d.). *Homepage: the FTW Transcriber*. Retrieved January 7, 2023, from the FTW Transcriber Corporation website: https://theftwtranscriber.com/

TOOLSADAY. (n.d.). *Text analysis, Similarity checker: Compare Text Online*. Retrieved January 17, 2023, from TOOLSADAY Web site: https://toolsaday.com/text-analysis/similarity-checker

United Nations Conference on Trade and Development. (2020). *Review of Maritime Transport 2020.* United Nations Conference on Trade and Development. Retrieved January 24, 2023

Verbeck, E. (2011, June). That dreaded 80 percent. *Seaways*, 24-2.

WORDCounter. (n.d.). *Character Counter*. Retrieved March 16, 2023, from WORDCounter Web site: https://wordcounter.net/character-count

Ziarati, R. (2006). Safety At Sea – Applying Pareto Analysis. *Proceedings of World Maritime Technology Conference (WMTC 06)*.

Ziarati, R., Ziarati, M., Bigland, O., & Acar, U. (2011). Communication and Practical Training Applied in Nautical Studies. *International Maritime English Conference.* Constanta, Romania.

9 Appendix

Interview 1

Interviewer: Ahmed Elhadi

Interviewee: Michael Stormbom, Chief Technical Officer at Lingsofty Oy

Date of interview: 6 March 2023

Interview 2

Interviewer: Ahmed Elhadi

Interviewee: Emin Nakilcioglu, Research Associate at Fraunhofer CML

Date of interview: 5 April 2023