Technische Universität Berlin

DOCUMENT BUILD DATE: October 7, 2019
DOCUMENT STATUS: Beta

Optimizing Agent Behavior and Minimizing User Cognitive Load for Mixed Human-Robot Teams

Master Thesis
am Fachgebiet Agententechnologien in betrieblichen Anwendungen und der Telekommunikation (AOT)
Prof. Dr.-Ing. habil. Sahin Albayrak
Fakultät IV Elektrotechnik und Informatik
Technische Universität Berlin

vorgelegt von
Christopher O'Hara

Betreuer:	Christopher-Eyk Hrabia,
Gutachter:	Prof. Dr.-Ing. habil. Sahin Albayrak
	Prof. Dr. Odej Kao

Christopher O'Hara
Matrikelnummer: 406037

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum Unterschrift

Abstract

DELETEME: An abstract is a teaser for your work. You try to convince a reader that it is worth reading your work. Normally, it makes to structure you abstract in this way:

- one paragraph on the motivation to your topic
- one paragraph on what approach you have chosen
- and one paragraph on your results which may be presented in comparison to other approaches that try to solve the same or a similar problem.

Abstract should not exceed one page (aubrey's opinion)

Zusammenfassung

DELETEME: translate to German to Englisch or vice-versa.

Acknowledgements

DELETEME: Thank you for the music, the songs I am singing

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Chapter 2

Motivation

Autonomous systems continue to improve at specialized tasks allowing for a higher quality of life for many individuals. Intelligent robotics have historically assumed tasks that would otherwise be arduous, vacuous, or even dangerous for humans. Specifically designed for distinct tasks, these machines do not grow tired over time and have efficiency/accuracy rates not typically achievable by humans (i.e., the manufacturing of vehicles or throughput of products on an assembly line). However, these topics are usually 'behind the scenes', as little-to-no human interaction is required. Recently, autonomous systems are proposed to coalesce and proliferate everyday activities via autonomous driving, product distribution, and tasks that require repetitive actions that do not require critical thinking.

Throughout the world, safety is a major concern for people and is a major factor of *Quality of Life* (QoL). In Germany alone, 23% of urban denizens report feeling unsafe due to crime in 2016 ([EU, 2016]). As such, there is a major need for improving the QoL for citizens, especially in urban areas. In general, police forces are responsible for the reduction of crime. Special forces units are called in whenever there are high-profile missions containing dangerous criminals (homicides, hostage situations, counterterrorism, etc.). In these situations, there is a high level of risk for the operatives that can result in permanent physical/mental damage or death. Another complication is that these operatives must wear additional protective gear. This gear blocks their ability to function ideally, as the total weight of gear can be upwards of 50kg and helmets/flak jackets block mobility. To combat these limitations, mixed human-robot teams can be assembled, sending in robots in to assess certain dangers (reconnaissance), structural damage, and even first contact with the perpetrator.

While certain aspects of autonomous systems currently prevent their abundance in urban environments (privacy concerns, resources, technological feasibility, etc.), there are additional niche areas that can immediately benefit from mixed human-robot teams for safety outside of police work. These stakeholders/areas/events include firefighters, disaster relief, item retrieval, contraband/explosives disarmament, and other environments in which the presence of a robot would reduce the risk of danger for human operators. In terms of interoperability and scalability, human-robot system design should find commonalities between these types of events and create a standardized framework.

Within the fields of autonomous systems and human-robot interaction, there is a lot of work to be completed. In the case of mixed human-robot teams, proper perception/cognition (for the robot) and distance (*Personal-Space Model*) for the user ([Torta et al., 2013]) is being analyzed with parametric models based on commonalities in user experience and comfort. Decision-making operations are currently in development for autonomous agents for reactive-adaptive hybrid behavior-based planning (*ROS Hybrid Behavior Planner* RHBP) ([Hrabia et al., 2017]). In fact, simple navigation and interaction with robots is not a trivial task that cannot be solved solely with intuition.

2.1 Project Scope

This project will be in direct relation to the needs of the Brandenburg *Spezialeinsatzkommandos* (SEK) (EN: Special Deployment Commandos), a German Special Forces group (Fig. 2.1). Their operations are often conducted in small, indoor environments where mobility and visibility are limited. As a special operation, counter-terrorism, and high-risk unit, they deal with the most critical tasks including hostage sieges, building raids, and surveillance ([Traumberuf, 2019]). Adding *Unmanned Aerial Vehicles* (UAVs) to their teams allows for a unique opportunity to minimize the danger of field operators during their missions. While certain military groups, like the *United States Marine Corps* (USMC), have been experimenting with drones (DE: Drohne) during operations, these have largely been outdoors and a safe distance from the target (Fig. 2.2)([Rogoway, 2018]). This project will have additional challenges in which the drone acts as a support member in a mixed human-robot team, providing information on target identification/classification, environmental hazards (smoke, fire), and structural layout.

Figure 2.1: Image of the Brandenburg SEK Force. Tactical gear prioritizes safety which negatively impacts vision, mobility, and cognitive load ([Moll, 2017]).

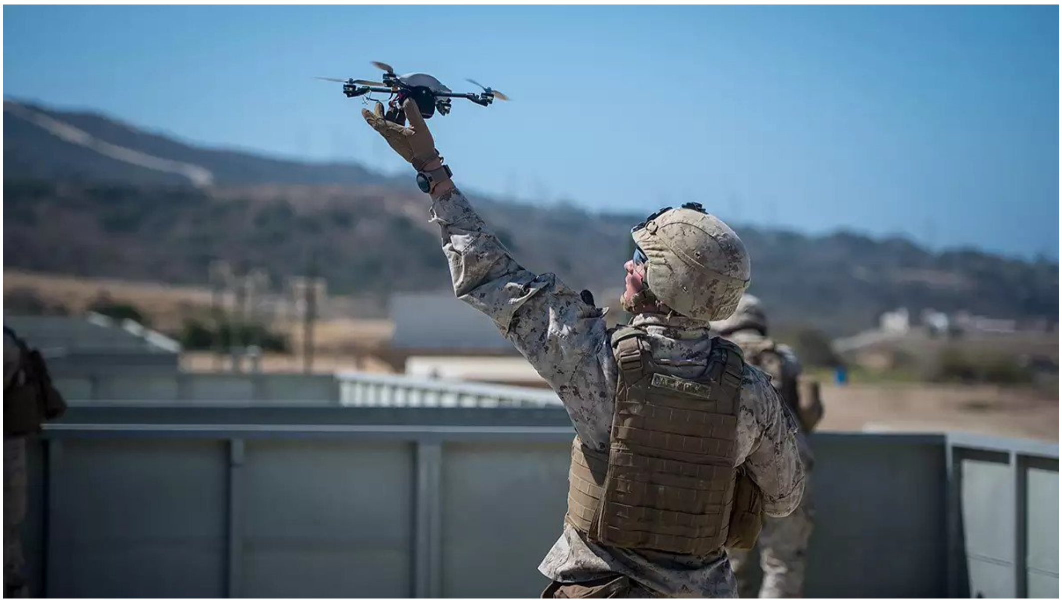

Figure 2.2: Image of USMC operative with a drone ([Rogoway, 2018]).

Chapter 3

Background

3.1 Speech Recognition

Controlling a drone via voice commands will require several steps. First, *Automatic Speech Recognition* (ASR) will need to be implemented. Since the drone will only be required to recognize and respond to a small set of words/phrases, creating a custom dictionary would be ideal (this will directly impact the memory size of the data set and can possibly result in more accurate interpretations). Next, the interpretations need to be processed and the drone's behavior needs to reflect the spoken dialog. Since the SEK is operating in dynamic and dangerous environments, this process needs to happen with high accuracy/precision and with as little delay as possible.

ASR has two main approaches in implementation: either using *Hidden Markov Models* (HMM) or *Deep-Learning* (DL). Each implementation has a number of trade-offs and the manner of the implementation determines which approach is, in some sense, optimal. HMMs are typically easier to understand and implement (fewer parameters) at the cost of accuracy ([Receveur and Fingscheidt, 2014]). Creating HMMs for ASR requires five steps: Feature Extraction, an Acoustic Model, a Lexicon Model, a Language Model and then a Decoder (usually the Forward-Backward or Viterbi algorithms) ([Maas, 2017]). HMMs are good for problems that have a small number of states (grammar, words) and could allow for the drone to perform entirely offline ([Ward, 2017]). There is already a large library of existing solutions and implementations utilizing the CMUSphinx API ([CMU-Sphinx, 2019]).

DL methods, usually with *Convolutional Neural Networks* (CNN) or Recurrent Neural Networks (RNN) (or a combination CNN-RNN encoder/decoder ([Wang et al., 2016])), allow for higher accuracy in results given proper parameter tuning and a large enough dataset ([Song, 2015]). DL implementations, on the other hand, will require that a network is trained with large amounts of data and the model be upload directly to the drone (or a microcontroller interface). In general, even using only an RNN will outperform an HMM ([Graves et al., 2006]). More research needs to be completed prior to selecting the best candidate for ASR, as it currently seems that a DL method would be the better candidate due to its higher accuracy. The current goal is to achieve high-performance results (and an excellent demonstration). However, the focus will be to create/implement/simulate a speech recognition model that is useful for evaluating user cognitive load.

Text-to-Speech and *Speech-to-Text* are two methods of machine translation. Intuitively, these technologies are mappings to shift from one domain to the other. These are common in modern ASR technologies (like Alexa and Siri) for providing the user with a confirmation of audio commands.

3.2 Mixed Reality User Interface

Once the drone can properly be given instructions via ASR, it is important that the drone is able to communicate back to the user in a meaningful manner. The main user-experience related criteria that are being pursued in this project is a 'seamless, natural integration of a drone into a mixed human-robot team.' This means that the drone needs to be able to communicate effectively (quickly, clearly, and without ambiguity). As such, there are very few approaches that would be sufficient in an environment that places a great deal of stress on the user. Haptic feedback (tactile vibrations) are likely not a good approach, as they can be easily misunderstood (and rely on the user being trained to interpret the message ([Cho and Proctor, 2003])). A speech response from the drone is also not ideal, as it requires that the user is primed for listening to the confirmation of the planned tasked ([Dyson, 2008]). During stress, the cognitive ability of the user will degrade ([Broadbent, 1958]). This means that long response might be forgotten, misunderstood, or not properly heard ([Engle, 2002]). Any of these events would require that the user ask the drone to repeat the flight plan. These are issues that exist even in human-human interactions, so it would be highly advantageous if a robotic system could mitigate or alleviate these issues.

An alternative to tactile and audio feedback is visual feedback. This will improve the response time of the user and mitigate errors due to multi-modal bottlenecks in information processing ([Sommer et al., 2001]). One approach would be to put LEDs on the drone that signal certain sequences of commands. Again, in this case, the operator needs to be trained to interpret these responses from the drone. Furthermore, once the drone is out of the line-of-sight, the drone will not be able to effectively communicate with visual commands. Therefore, a UI within a Mixed Reality device (i.e., HoloLens) can also be utilized. The UI can be designed in a mixed-reality format that displays, with written words or intuitive icons, the flight/task plan of the drone. It is hypothesized that this will mitigate the forgetfulness of the user due to the high cognitive load. The UI could be designed in a manner that will not impede the vision of the user. This allows the user to maintain situational awareness and analyze the feedback when deemed safe. Regardless, both audio and visual solutions need to be approached as often times theory and reality do not match in real-world implementations.

3.3 Robot Behavior Planning

In mixed human-robot teams, predictable and reliable agent behavior is crucial. [Hrabia et al., 2017] have created a framework that allows for multi-agent environments with adjustable parameters such as pose, distance, and movements. When considering having a drone (or another robot agent) adjust behavior depending on both the user and the environment (both dynamic), a flexible approach to decision making and planning is desired. This would allow for the drone to change is behavior in real-time without compromising the mobility/visibility of the operator or negatively interacting with the environment (i.e., crashing into a wall).

The *ROS Hybrid Behavior Planner* (RHBP) seeks to meet these requirements with a hybrid reactive-adaptive behavior-based planner. This is a major improvement over traditional approaches (conditional statements) that do not have any weights, probabilistic characteristics or allow for (potential) reinforcement learning models.

3.4 Cognitive Load Theory and Testing

Many different tests have been created to evaluate the cognitive performance of individuals during various conditions. *Situational Awareness* is limited by available resources in memory capacity and computation during decision-making ([Endsley, 1995]). When resources are limited or have been depleted, users begin to make errors in accuracy or decision making as a result. Cognitive load (and SA) can be evaluated with two types of tests: qualitative and quantitative. A common qualitative test is the NASA Task Load Index (NASA-TXL) which allows users to access their *perceived* stress and workload levels while completing other tasks. Another traditional test is based on Cognitive Load Theory (CLT) in which subjects report their perceived performance after tasks are given for problem-solving or spit-attention ([Chandler and Sweller, 1991]).

For qualitative experiments, attention, working memory, and response time are common metrics. The *Operation Span Task* (OSPAN) is used to measure the *Working Memory* capacity of users ([Turner and Engle, 1989]). This is completed by giving the user an item to remember and then interjecting with random simple mathematics problems, after which they are asked to recall the item. In the case of drone operations, a set of directions could be provided as the ‘item.’ The *Attentional Blink* test is used to measure the response time and accuracy of a user when two items are displayed for brief periods of time with a short (variable) delay between objects ([Nieuwenstein et al., 2009]).

NASA has employed the *Psychomotor Vigilance Task* to measure the cognitive performance of astronauts during space missions (NASA Extreme Environment Mission Operation NEEMO ([Dinges, 2019a])) and aboard the International Space Station. This test measures the response time for the user to press a keyboard after an image has been displayed on a screen (faster is better).

Chapter 4

My First Main Part

4.1 Objectives

Note: each RQ will be approached and completed with the mindset that this is not a final, shippable product, i.e., this is a prototype/mock-up. In reality, the foundation and framework need to be available in the form of a robust 'proof-of-concept' that can be adaptable for specific needs of the stakeholders that follow (e.g., the SEK will likely want to change specific details as well as future researchers within the DAI-Labor).

Each research question will be broken down into work packages. Each implementation will be given clear objectives and criteria for evaluation. Some work packages will have a slight overlap related to tuning, implementation, and iterative design. The end results are intended for the interoperability between distributed systems, applicable to many projects.

4.1.1 RQ1: Human-Robot Communication

RQ1: How should the drone send confirmations to the operator to minimize ambiguity, human errors, and temporal descrepancies?

After viewing the field demonstration and discussion with the SEK, it is evident controlling the drone via voice commands is preferred. The SEK operatives primarily utilize speech (through headsets) for communication (local and global). While gestures, positioning, and other non-verbal communication methods are employed (i.e., patting and tapping) this are based on proximity (locality) and the situation. Furthermore, as local communication is based on quick confirmations, it does not appear to impact the overall goal. As such, the drone will still have its core mission criteria that are independent of these non-verbal interactions.

Once the drone has received its orders, it should begin to complete its tasks. However, it is currently not clear how the drone will communicate with the operator that it has correctly interpreted the task. Consider the following task given to a drone: 'go forward to an open doorway, enter the room, scan the area, proceed into any additional room on the right side.' It is possible that the drone correctly begins its task and enters the first room but did not properly interpret the sequence afterward. If the drone is able to confirm the sequence with the operator, proper navigation can be ensured (at least on the command level). Furthermore, it would be ideal to give a drone a sequence of commands with a 'wait' feature, i.e., `sleep()`. The operator could properly confirm the sequence, make any necessary changes, and then send an 'engage' command. The drone should also be able to send confirmations once each individual command has been completed (like a check-list) if desired.

In designing *Natural User Interfaces* (NUIs), research has shown that gesture-based interfaces are less natural for communication. Due to the high criticality of missions and adherence to safety, any additional ambiguity related to NUIs would impact mission performance. Audial and visual interfaces can decrease ambiguity if properly designed. The first logical step is to determine which of these two interfaces are preferred prior to detailed implementation. Therefore, a user study needs to be conducted evaluating preference, subjective load, and qualitative factors.

4.1.2 RQ2: Operator Cognitive Load

RQ2: What are the negative effects (response time delay, accuracy, situational awareness) on the operator's cognitive load during operation?

Automatic Speech Recognition (ASR) models are not perfect. The middleware might not properly understand the commands of the user which greatly impacts the mission's objectives due to a loss of time for correction and operator frustration/distraction. A user study will be needed in which the user completes some simple tasks simultaneously as validating the drone's confirmation. The concept here is that a drone could report to the operator incorrect information which could lead to mission failure and/or compromise the operator/plan. For quantitative evaluation, several tests from the cognitive science field have been employed to measure *Working Memory* (WM) capacity (is the operator able to properly remember the sent commands when under pressure?) and *Response Time* during events requiring a high cognitive load and split attention. The results and analysis will infer user's situational awareness and confirmation accuracy. The user could be asked to give commands to the drone and listen for the confirmation while taking the test. Quantitative results would be compared between a baseline (test only), test while receiving audial confirmations (headset), and tests while receiving visual confirmation (AR).

4.2 Proposed Solution

Based on the previous sections, the goal is to enable a drone to accept user commands, accurately follow the commands, and send a confirmation to the user (for review). Prior to selecting the form in which the drone should send confirmations, qualitative (user preference) and quantitative (reaction times, accuracy of memory) need to be derived from user study results. The simulation environment will be based on the previous work at DAI-Labor ([Hrabia et al., 2017]). A custom library/dictionary will be created for ASR and validated in ROS. ROS will then communicate these commands to the drone (SST) and the drone's behavior should follow correctly. It is planned to create a UI for visual representations of the confirmation (a text-based or icon-based visualization) depending on the Mixed-Reality device (i.e., Unity 3D for HoloLens).

4.2.1 Milestone I

RQ1 and RQ2 will be approached simultaneously. The first aspect is to implement a library/dictionary for *Automated Speech Recognition* (ASR) that can allow for commands to be sent to a drone (either simulated or actually sent to a drone via ROS). Next, a Speech-to-Text (STT) library could be designed for machine translation into ROS (ideal of verification of commands). In the physical implementation of a drone, the STT commands would be sent to the robot. For cognitive evaluation alone, it might be better to focus on a Text-to-Speech (TTS) model. Utilizing a terminal, TTS commands could be manually typed and sent to the user's device to evaluate their hit/miss rate for confirming the drone's response. This would mitigate any errors that would be caused by utilizing ASR (control variable). TTS commands would be sent to the user's headset during a cognitive test. For an MR solution, STT would allow for commands to be translated and sent to the device's UI in the form of words or icons that represent words (i.e., a right arrow for 'right').

It is intended to use the NASA-TXL (or similar) for measuring perceived cognitive load while asking users to make a preference for receiving audial or visual confirmations. For quantitative tasks, it is intended to use the Inquisit Laboratory program with the PVT test. Both tests are regularly used by NASA to evaluate astronauts during space missions, furthermore, NASA quotes: "The PVT Self Test has wide application to any group that must operate remotely at high levels of alertness, such as first responders, Homeland Security personnel, flight crews, special military operations, police, and firefighters ([Dinges, 2019b])." This makes the test ideal in evaluating subjects for high cognitive load tasks. The current plan is to evaluate three use cases for the test: one without any drone commands (baseline), one with audial drone confirmations, and one with visual drone confirmations. The goal is to have the users complete the tasks of the PVT while measuring their response times and accuracy (automatic with Inquisit Laboratory ([Millisecond, 2019])) and documenting their accuracy for confirming correct drone command sequences (hit).

4.2.2 Milestone II

After qualitative and quantitative results have been analyzed and weighed, a virtual implementation can be thoroughly constructed in ROS (with an RHBP model). The task will be to ensure that drone behavior matches the input commands and the output is properly sent to the user's device. Additionally, the drone will operate within the selected range/bounds based on the operational criteria based on SEK operatives, as well as an interoperable model that can be used for general projects (a distributed model). Finally, once the behavior has been validated and verified in ROS, it is intended to provide a live demonstration with the drone responding to the user commands, following intended behavior, and reporting back to the user's device.

4.2.3 Project Management

Typical project management aspects will also follow. A GitLab account will be created via TU Berlin and Scrum-style weekly meetings will be conducted with Christopher-Eyk Hrabia. This will help to ensure timeliness and proper communication. Deviations are likely to happen (i.e., switching from one framework to another due to compatibility or performance issues) but this will be mitigated as much as possible with proper literature review and stand-alone implementations.

4.3 Work Packages

The project will be broken down into several work packages. Figure **??** gives the timeline for task completion. This section proposes some of the current approaches and architectures.

4.3.1 Research

Research has been conducted, thus far, by first evaluating the goal of the project. This has made the approach easier by evaluating how other researchers have approached this issue (from classical methods to state-of-the-art). Next, it was critical to ensure the alignment of the needs of the SEK. This pruned a lot of approaches and methods. However, more iterations within the research cycle need to continue in order to meet the high-performance metrics while maintaining feasibility.

4.3.2 Implementation Process

It is planned to make a library based on the commands that SEK operatives would use during operations. Such words might include 'forward', 'right', and 'stop'. Furthermore, temporal and spatial commands would be ideal ('forward for five seconds', 'left two meters'). However, spatial commands need a proper translation from the user's perspective to the drone's perspective. If possible (and in existence) it would be good to allow for a command that switches to autonomous behavior ('exploration/scan mode'). This will depend on what the drone is already capable of (or added as part of this project by pruning another research question). Each utterance (word) will need to be recorded at least 200 times. This will help provide enough data for a DL implementation with TensorFlow (see below). If this is insufficient, more recordings will be created. For HMM models, it is not clear how much training data is needed currently. The resulting model can be tested (if required) with users but it might be better to treat the proposed user testing with a 'simulated' environment (i.e., manually sending congruent/incongruent command sequences via TTS or text-based inputs).

Until now, a large amount of time was spent to ensure that the project was feasible. ROS is intended to be the core middleware for program communication. Many considerations were given based on the (real) environment that the drone would operate in. For example, Cloud computing for ASR allows for the highest accuracy using Google's STT framework ([Google, 2019b]). However, Cloud interfaces might not be plausible and soldiers might not be in an environment that can directly interact with the Cloud.

Currently, the plan is to create an independent library and either utilize DL via TensorFlow ([TensorFlow, 2019]) based on Small-footprint Keyword Spotting ([Chen et al., 2014]) or use a HMM model with PocketSphinx and a custom dictionary. TensorFlow can easily integrate with ROS and is relatively easy to adjust. Furthermore, the TensorFlow implementation allows for continuous speech input and the trained models can be exported to Android (good for in-person demonstrations). Finally, results can be visualized using TensorBoard, an aesthetic and informative tool. In terms of an HMM approach, many researchers have used CMU-Sphinx (PocketSphinx) for ASR. Achieving the highest accuracy possible is desired, so the final implementation may require the DL/TensorFlow approach as HMM can have variable accuracy as the end result (which is undesirable for safety-critical missions) ([Telmem and Ghanou, 2018]), ([Plonkowski, 2018]).

In the event there are issues with the speech recognition in ROS, it can be bypassed and completed in Unity 3D (worst-case scenario) ([Unity, 2019]). Unity can act as middleware, though it often behaves slower. This would only be in the case of a demonstration that was also utilizing the HoloLens (as is done by Siemens ([Siemens, 2019])). However, a HoloLens might not be available and another device (which will likely be lighter and better for cognitive load) will need to be used. If it does not have an interface with Unity, another solution will be employed. Another option is to directly employ the Google Cloud Speech API in ROS. Again, based on the assumptions I have proposed, this is not ideal and would be left only for the sake of demonstration (i.e., for the SEK with a very constrained deadline) ([Google, 2019a]).

The UI will be created using Unity 3D (HoloLens) or similar software (third-party AR/MR device). This allows for some visualizations and initial feedback even without having a device on hand. In fact, it could be that this lays the framework for implementation (as a proof-of-concept) in the event that a HoloLens is unobtainable.

Chapter 5

My Second Main Part

Chapter 6

Evaluation

DELETEME: The evaluation chapter is one of the most important chapters of your work. Here, you will prove usability/efficiency of your approach by presenting and interpreting your results. You should discuss your results and interprete them, if possible. Drawing conclusions on the results will be one important point that your estimators will refer to when grading your work.

6.1 Results

6.2 Discussions

Chapter 7

Conclusion and Future Work

7.1 Summary

DELETEME: put a plain summary of your work here. Summaries should be made of each Chapter beginning with Chapter 2 and ending with you evaluation. Just write down what you did and describe the corresponding results without reflecting on them.

7.2 Conclusion

DELETEME: do not summarize here. Reflect on the results that you have achieved. What might be the reasons and meanings of these? Did you make improvements in comparison to the state of the art? What are the good points about your results and work? What are the drawbacks?

7.3 Future Work

DELETEME: Regarding your results - which problems did you not solve? Which questions are still open? Which new questions arised? How should someone / would you continue working in your thesis field basing on your results?

Bibliography

[Broadbent, 1958] Broadbent, D. (1958). *Perception and Communication*. Pergamon Press.

[Chandler and Sweller, 1991] Chandler, P. and Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Faculty of Education - Papers*, 8.

[Chen et al., 2014] Chen, G., Parada, C., and Heigold, G. . (2014). Small-footprint keyword spotting using deep neural networks. pages 4087–4091.

[Cho and Proctor, 2003] Cho, Y. S. and Proctor, R. W. (2003). Stimulus and Response Representations Underlying Orthogonal Stimulus-Response Compatibility Effects. *Psychonomic Bulletin & Review*, 10(1):45–73.

[CMU-Sphinx, 2019] CMU-Sphinx (2019). ROS-Pocketsphinx.

[Dinges, 2019a] Dinges, D. F. (2019a). Experiment Summary Reaction Self Test.

[Dinges, 2019b] Dinges, D. F. (2019b). Psychomotor Vigilance Self Test on ISS.

[Dyson, 2008] Dyson, P. (2008). *Cognitive Psychology*. British medical bulletin. Academic Press.

[Endsley, 1995] Endsley, M. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37:32–64.

[Engle, 2002] Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1):19–23.

[EU, 2016] EU (2016). Quality of Life Indicators - Economic and Physical Safety.

[Google, 2019a] Google (2019a). Google cloud speech for ros.

[Google, 2019b] Google (2019b). Google Cloud Speech-to-Text API.

[Graves et al., 2006] Graves, A., Fernández, S., Gomez, F., and Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural 'networks. volume 2006, pages 369–376.

[Hrabia et al., 2017] Hrabia, C., Wypler, S., and Albayrak, S. (2017). Towards Goal-Driven Behaviour Control of Multi-Robot Systems. In *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*, pages 166–173.

[Maas, 2017] Maas, A. (2017). Spoken Language Processing-ASR: HMMs, Forward, Viterbi.

[Markus.Porto - Uni Giessen,] Markus.Porto - Uni Giessen. Kochbuch für latex. `http://www.uni-giessen.de/hrz/tex/cookbook/cookbook.html` (3. April 2007).

[Millisecond, 2019] Millisecond (2019). Inquisit Lab.

[Moll, 2017] Moll, T. (2017). SEK Brandenburg Erhielt Achleitner HMV-Survivor.

[Nieuwenstein et al., 2009] Nieuwenstein, M. R., Potter, M. C., and Theeuwes, J. (2009). Unmasking the Attentional Blink. *Journal of Experimental Psychology, Human Perception and Performance*, 35(1):159–169.

[Oberheide et al., 2008] Oberheide, J., Cooke, E., and Jahanian, F. (2008). Cloudav: N-version antivirus in the network cloud. In *Proceedings of the 17th USENIX Security Symposium (Security'08)*, San Jose, CA.

[Plonkowski, 2018] Plonkowski, M. (2018). Using full covariance matrix for cmu sphinx-iii speech recognition system. *PRZEGLAD ELEKTROTECHNICZNY*, 1:104–106.

[Receveur and Fingscheidt, 2014] Receveur, S. and Fingscheidt, T. (2014). On likelihood histogram equalization for multimodal automatic speech recognition.

[Rogoway, 2018] Rogoway, T. (2018). Marine Corps Considers Deploying Small Drones at Squad Level.

[Siemens, 2019] Siemens (2019). Siemens: Ros sharp for unity.

[Sommer et al., 2001] Sommer, W., Leuthold, H., and Schubert, T. (2001). Multiple Bottlenecks in Information Processing? An Electrophysiological Examination. *Psychonomic bulletin & review*, 8 1:81–8.

[Song, 2015] Song, W. (2015). End-to-end deep neural network for automatic speech recognition. Stanford University.

[Telmem and Ghanou, 2018] Telmem, M. and Ghanou, Y. (2018). Estimation of the optimal hmm parameters for amazigh speech recognition system using cmu-sphinx. *Procedia Computer Science — Proceedings of the First International Conference on Intelligent Computing in Data Sciences*, 127:92–101.

[TensorFlow, 2019] TensorFlow (2019). Tensorflow: Audio recognition.

[Torta et al., 2013] Torta, E., Cuijpers, R., and Juola, J. (2013). Design of a Parametric Model of Personal Space for Robotic Social Navigation. *International Journal of Social Robotics*, 5(3):357–365.

[Traumberuf, 2019] Traumberuf (2019). Berufsbild-Polizist Spezialeinsatzkommando.

[Turner and Engle, 1989] Turner, M. L. and Engle, R. W. (1989). Is Working Memory Capacity Task Dependent? *Journal of Memory and Language*, 28.

[Unity, 2019] Unity (2019). Voice input in unity.

[Wang et al., 2016] Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., and Xu., W. (2016). CNN-RNN: A Unified Framework for Multi-label Image Classification. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2285–2294.

[Ward, 2017] Ward, W. (2017). Hidden Markov Models in Speech Recognition.

Appendices

DELETEME: everything that does not fit into your work, e.g. a 5 page table that breaks the reading flow, should be placed here

Appendix A: Abbreviations

ASR	Automatic Speech Recognition (Automatische Spracherkennung)
CNN	Convolutional Neural Network
DL	Deep-Learning (Tiefes Lernen)
HMM	Hidden Markov Model (verdecktes Markowmodell)
NEEMO	NASA Extreme Environment Mission Operation
OSPAN	Operation Span Task (Operationsspanne Aufgabe)
RNN	Recurrent Neural Network (Rekurrentes neuronales Netz)
LED	Light Emitting Diode (lichtemittierende Diode)
LSB	Least Significant Bit
MD5	Message Digest (Kryptographisches Fingerabdruckverfahren)
MPEG	Moving Picture Experts Group (Video- einschließlich Audiokompression)
MP3	MPEG-1 Audio Layer 3 (Audiokompressionformat)
PACS	Picture Archiving and Communication Systems
PNG	Portable Network Graphics (Grafikformat)
RSA	Rivest, Shamir, Adleman (asymmetrisches Verschlüsselungsverfahren)
SHA1	Security Hash Algorithm (Kryptographisches Fingerabdruckverfahren)
WAV	Waveform Audio Format (Audiokompressionsformat von Microsoft)

Appendix B: LaTeX Help

How to Use This Template

- Remove all of my text which is mostly labeled with DELETEME
- Change the information in the 00a_title_page.tex file
- Use the information written in this section
- Ask you supervizor to help you
- If I am not your supervizor and noone else can help you, write me an email (aubrey.schmidt@dai-labor.de)

Citations

Citing is one of the essential points you need to do in you thesis. Statements not basing on results of your own research[1] not being cited represent a breach on the rules of scientific working. Therefore, you every statement needs to be cited basing on information that other people can cross-check. A common way of citing in technical papers is:

- Oberheide et al. [Oberheide et al., 2008] state that the average time for an anti-virus enginge to be updated with a signature to detect an unknown threat is 48 days.

Note: et al. is used when the paper was written by more than two people. Check the code of this section to learn how to cite from a technical perspective.

Note: you can change the citation style in the `thesis.tex` file, e.g. to harvard style citations. Instructions on this can also be found in this file.

You should not cite anything that can be changed, e.g. it is not that good citing web pages since they might get updated changing the cited content. There are no clear quality measures on citing sources but aubrey believes that the following list is true for several cases, starting with highest quality:

1. Journal article or book
2. Conference paper
3. Workshop paper
4. Technical report

[1] in what ever context

5. Master thesis
6. Bachelor thesis
7. General Web reference

There might be workshop papers that have a higher quality than some journal papers. Therefore this list only gives you a hint on possible quality measures. Another measure can be whether a paper was indexed by ACM/IEEE, although this is not a strong indicator.

Finding and Handling Citation Sources

Following ressources are required for finding and handling articles, books, papers and sources.

- your primary resource will be `http://scholar.google.com`
- `http://www.google.com` might also be used
- `wikipedia.com` can be a good start for finding relevant papers on your topic
- you should download and install JabRef or a similar tool `http://jabref.sourceforge.net/`
- you should point JabRef to the mybib.bib file
- you should immediately enter a relevant paper to JabRef, additionally, you should write a short summary on it; else, you will do this work at least twice.

General Advices

- Do not take care of design, LaTeXwill do this for you. If you still feel that you need to take care of this, do this when you have finished writing, else you will end up in a lot of double and triple work.
- LaTeXwill do exactly that you will tell it to do. If you have problems with this, go for google or ask you supervizor
- use labels in order to be able to reference to chapters, section, subsections, figures, tables, etc. ...

General Commands

- check `http://en.wikibooks.org/wiki/LaTeX`
- check `http://www.uni-giessen.de/hrz/tex/cookbook/cookbook.html` German

Please also check the following source [Markus.Porto - Uni Giessen,].

Code

This section shows you how to get your code into a LaTeXdocument. See code for options.

```
class Beispiel{

  public static void main(String args[]){

    System.out.println("Hello␣World");

  }

}
```

```
class Beispiel{

  public static void main(String args[]){

    System.out.println("Hello World");

  }

}
```

Listing 7.1: Example code is presented here

Figures

This section describes how to include images to your document. Information was taken from `http://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions`, visited on 05/08/2011. Please make sure to use original vector graphics as basis since image quality might be used as weak indicator for thesis quality. For this, try to find find files in `.SVG` or `.PDF` format. Exporting a `.PNG` or `.JPG` to `.PDF` will not work since data was already lost while exporting it to these formats. This is the case for most Web graphics. Wikipedia startet entering most in images in `.SVG` which easily can be transformed to `.PDF`, but please do not forget proper citations.

Figure 7.1: Including an image; in this case a PDF. Please note that the caption is placed below the image.

Figure 7.2: See code for caption options: this is a long caption which is printed in the Text. Additionally, image size was increased

(a) Small (b) Large (c) Medium

Figure 7.3: Placing images side by side using the subfig package. Space between the images can be adjusted.

Tables

Here, you will find some example tables.The tables were taken from `http://en.wikibooks.org/wiki/LaTeX/Tables`, visited on 05/08/2011. Table environment was added plus caption and label. For code, check `__help/latex_hinweise.tex`.

Table 7.1: Simple table using vertical lines. Note that the caption is always above the table! Please check code for finding the right place for the table label.

1	2	3
4	5	6
7	8	9

Table 7.2: Table using vertical and horizontal lines

7C0	hexadecimal
3700	octal
11111000000	binary
1984	decimal

Table 7.3: Table with column width specification on last column

Day	Min Temp	Max Temp	Summary
Monday	11C	22C	A clear day with lots of sunshine. However, the strong breeze will bring down the temperatures.
Tuesday	9C	19C	Cloudy with rain, across many northern regions. Clear spells across most of Scotland and Northern Ireland, but rain reaching the far northwest.

Table 7.4: Table using multi-column and multirow

Team sheet		
Goalkeeper	GK	Paul Robinson
Defenders	LB	Lucus Radebe
	DC	Michael Duberry
	DC	Dominic Matteo
	RB	Didier Domi
Midfielders	MC	David Batty
	MC	Eirik Bakke
	MC	Jody Morris
Forward	FW	Jamie McMaster
Strikers	ST	Alan Smith
	ST	Mark Viduka