AALBORG UNIVERSITY
COPENHAGEN

Semester: 4th, Spring 2017

Title: Applying Cognitive Technologies and Emulating Interpersonal Communication in Affective Tutoring Systems

Aalborg University Copenhagen
A.C. Meyers Vænge 15
2450 København SV

Semester Coordinator: Henning Olesen

Secretary: Maiken Keller

Project Period: 1/2/2017-30/6/2017

Semester Theme: Master Thesis

Supervisor(s): Anders Henten

Project group no.: 4BUS 4.26

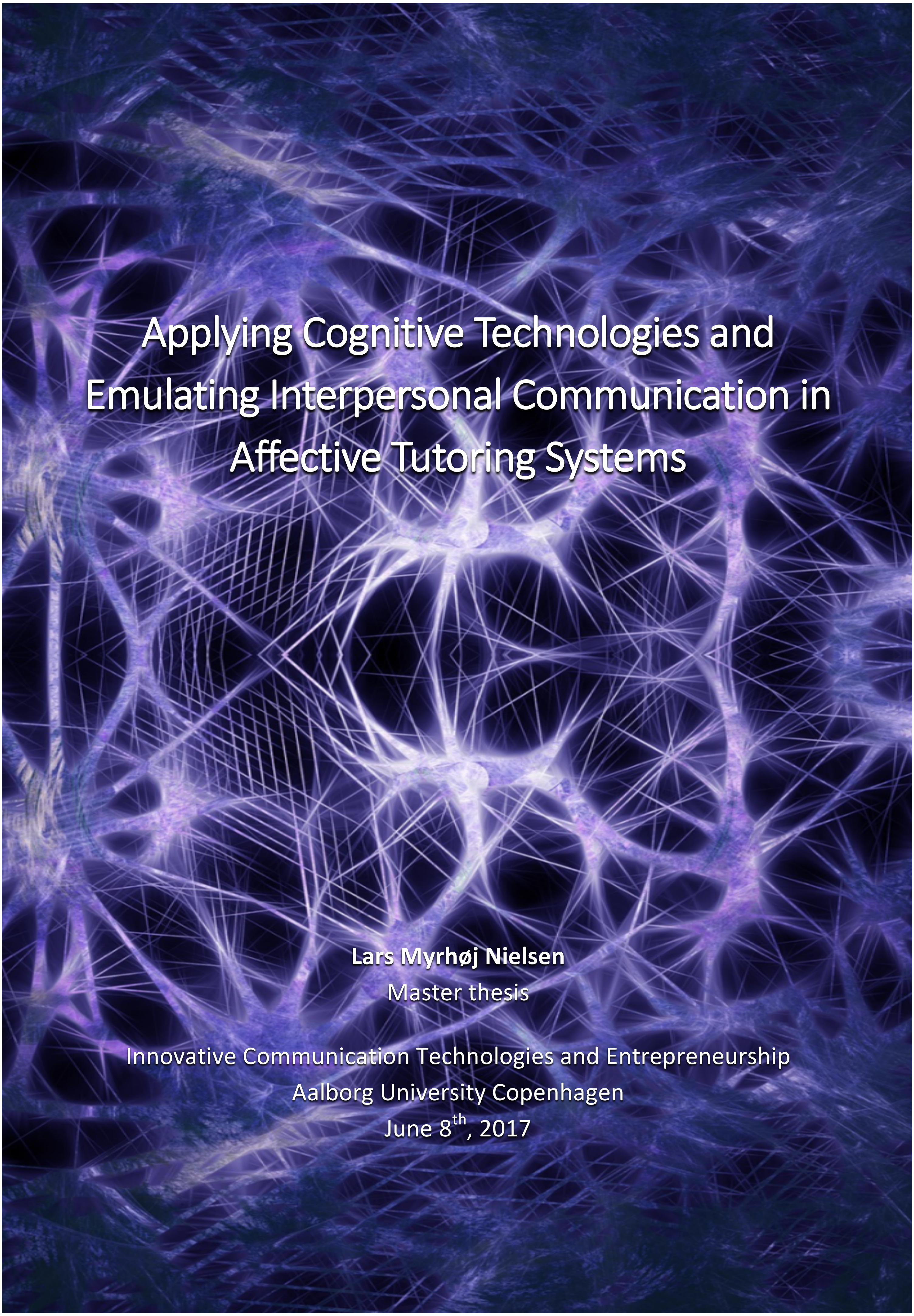
Members
(do not write CPR.nr.):
Lars Myrhøj Nielsen

Pages: 91
Finished: 8/6/2017

Abstract:
It is well-documented that Intelligent Tutoring Systems (ITS) have a positive effect on the learning outcome across all educational levels, compared to most traditional learning methods. ITS are able to provide individualized tutoring, based on individual traits of the learner. However, research shows that ITS are still inferior to both small-group tutoring and individual tutoring. This is mainly a result of the teachers' ability to take the affective/emotional state of the individual student into consideration when communicating.

This project investigates how cognitive and affective computing technologies can be used to emulate interpersonal communication, and thereby improve the effectiveness of the tutoring systems. For this purpose, a conceptual solution will be presented, based on theory on interpersonal communication and human emotion, and state of the art cognitive and affective technologies.

The findings from the project proved that it is possible to emulate the interpersonal communication process in an ATS system, to a certain degree. State of the art cognitive technologies enable processing of complex multimodal signals, similar to how communication signals are processed in the human brain, but their overall capabilities do not yet compare to the capabilities of the human brain.



Applying Cognitive Technologies and Emulating Interpersonal Communication in Affective Tutoring Systems

Lars Myrhøj Nielsen

Master thesis

Innovative Communication Technologies and Entrepreneurship
Aalborg University Copenhagen

June 8th, 2017

Table of Contents

1. Introduction

Over the past decade, the evolution of information technology has opened a new world of possibilities within the technology sector. Along with vastly improved computer algorithms, the computational power has now reached a level of sophistication, that the computers now appear to have the capability of actually thinking. They are able to understand complex information, and learn, reason, and act upon that information. Some systems can take contextual and environmental factors into account, and even adapt and evolve over time (Raghavan et al., 2016). As a result of these advancements we are now considered to be entering a third era of computing, the cognitive era (Kelly & Hamm, 2013). In 2011, IBM's cognitive supercomputer, Watson, beat two human champions in the game show *Jeopardy* (Jackson, 2011). Five years later, in 2016, DeepBlue's AlphaGo defeated the World's best player of the game *Go,* which is widely considered the most complex board game invented by mankind (Hassabis, 2016). These are just two examples of what is possible with cognitive computing, but it has the potential to revolutionise technology across all sectors (Pereira et al., 2015).

While technology has been used to support learning for decades, adaptive and interactive programs have now been developed, which can can provide advanced individualised tutoring. These systems are known as Intelligent Tutoring Systems (ITS), and are capable of modelling and acting upon learners' individual traits, such as psychological states, learning characteristics, needs, and pace of learning. By adapting to the individual student, ITS are able to guide students through different steps in problem solving, and offers a number of advantages over traditional classroom teaching. Numerous researchers have investigated the effect ITS has on the learning outcome, and found that using ITS is generally associated with greatly increased learning outcomes, compared to teacher-led large-group instructions, non-ITS computer-based instruction, and textbooks or workbooks (Kulik & Fletcher, 2015; Ma et al., 2014; Steenbergen-Hu & Cooper, 2014; VanLehn, 2011). This research did, however, also conclude that small-group and individual human tutoring was the most effective learning method, thus highlighting the important roles of teachers and pedagogic in learning. The reason for the superiority of human tutoring, is a result of the teachers' abilities to identify and respond to the student's affective state (Mao & Li, 2010). A student's emotional state have great impact on e.g. information processing, decision-making, and creative problem solving, and subsequently affect the learning effectiveness (Erez & Isen, 2002).

Thus, the key to enhancing the capabilities and effectiveness of ITSs, it widely considered to be enabling the systems to adapt their communication to the student's emotional state, in the same way human teachers do (Sarrafzadeh et al., 2008). Such systems are known as Affective Tutoring Systems (ATS). By detecting emotional states, responding to these states, generating relevant tutoring strategies, and communicating with affect, ATS aim to help regulate the student's emotional state, and achieve better learning outcomes. ATS employs so called affective computing, a subset of cognitive computing, which refers to computing related to, arising from, or influencing emotions (Picard, 1997). The overarching end-goal of affective computing is for humans to be able to interact and interface with machines as richly as we interact with each other (Cowie, 2005). Affective computing utilises technologies such as visual recognition, speech recognition, and physiological signals, to recognise and interpret human emotions. The educational sector, however, has not yet been revolutionised by cognitive and affective computing solutions, but it is an area in which it has a great potential use.

This project aims to investigate how cognitive and affective computing can be used to improve the interpersonal communication between teachers and students, and subsequently affect the effectiveness of ATS and the learning outcome.

1.1 Problem definition

Within human communication, cognitive and affective technologies, and educational learning, there are plenty of interesting topics to research. Below is a presentation of the focus of this project, in form of the research question, and a project delimitation.

Research question

Based on the preliminary research presented above, this project aims to answer the following research question:

> *How can cognitive computing technologies be used in Affective Tutoring Systems to emulate interpersonal communication between students and teachers?*

Project delimitation

Due to time constrains and limited resources, there are certain limitations as to what this project will include. The emphasis of this project is on interpersonal communication and the underlying communication processes, and how humans communicate. The aim is to investigate how this can be emulated within an ATS, however, the area of Human-Computer Interaction (HCI) will only be lightly touched upon, and not investigated thoroughly.

Another relevant topic which is only discussed superficially, is that of learning processes and learning theories. These research areas could prove beneficial in later stages or future work of the project, but they will not be considered at this stage.

With a focus on the research on how cognitive computing can be used within an ATS, it is not within the scope of this project to develop a functional ATS. The focus here is on the conceptual development, however, cognitive technologies which could be used to build a functional ATS, will be discussed.

1.2 Reading guide

In this introductory chapter, preliminary research within the areas of cognitive computing and ITS was used to define and justify the research question. Chapter 2 presents the methodological approach of the research. It describes the different stages of the project, the main components of the report, and the topics that will be addressed, to properly evaluate the research question.

In chapter 3, additional research will be made within the fields of ITS and ATS, by presenting the architectural components of such systems, as well as a more detailed description of the benefits of using tutoring systems. Furthermore, interviews will be made with a number of university students, to support the research findings.

In chapter 4, the first general topic of interest is addressed, that of interpersonal communication. Here, the most interesting aspects of interpersonal communication will be investigated, with the intention of constructing a model which can be used for further analysis later in the project. Chapter 5 addresses the second major element of the project, cognitive and affective computing. This includes a more thorough description of the two, and some of the underlying processes which enables cognitive computing. Furthermore, the current state of the art will be investigated and presented.

The main analysis will be presented in chapter 6. An ATS model will be presented, which will be discussed in relation to the interpersonal communication model constructed in chapter 4. To aid to this discussion, and add a more practical dimension, a prototype of an ATS will be developed. Both these aspects of the discussion will be addressed from the student's perspective as well, with follow-up interviews with the students.

Chapter 7 will discuss the most evident results from the analysis, in relation to the viability hereof, and address any shortcomings that might surface. The main conclusions will be presented in chapter 8, including a summary of the findings from the project, and an answer to the research question. Finally, chapter 9 will discuss ideas for future work on the project, based on the conclusions and the project as a whole.

2. Methodology

To gain a detailed understanding of ITS and ATS, additional research was made within these two fields. This researched aimed to elaborate upon what such systems entail, how they are built, what their purposes and benefits are, and how the outcome of using them actually compares to traditional forms of learning. Four students from Aalborg University Copenhagen were then interviewed, to gain a first hand impression of their learning experiences and preferences. It was also of interest to investigate the relationship between the students and their teachers, evaluate their individual preferences for interpersonal interaction and feedback, and to explore which specific elements and functionalities to include in an ATS. The students were interviewed individually, based on a set of pre-defined open-ended questions.

Before analysing how computers can emulate human communication, it is relevant to fully understand the different aspects of interpersonal communication. Thus, topics such as communication elements and processes, verbal and nonverbal communication signals, and different communication theories were investigated. Additional emphasis was then put on the effect and functionality of emotions within interpersonal communications, to better understand how emotion signals and emotional states affect the communication process. Based on research within these areas, an interpersonal communication model was constructed, illustrating the most relevant components.

The second major topic of interest in this project was that of cognitive computing. This topic was thoroughly researched, in regards to both the processes behind cognitive computing, its current and potential use, and the current state of the art of cognitive technologies. Due to the effect emotions has on both learning and interpersonal communication, affective computing was also investigated.

Based on this extensive research on both ITS and ATS, interpersonal communication and capabilities of cognitive technologies, a conceptual model of an ATS were proposed. This model took into consideration all aspects of the theoretical research. Based on this concept, a second model was proposed, providing a concrete example of how existing cognitive services could be used to implement a small-scale ATS.

The ATS model were then evaluated from two different perspectives. First of all, it was analysed according to the previously constructed interpersonal communication model, and discussed in regards to which extent the system can emulate individual human tutoring. Secondly, the conceptual model was discussed with the same four university students that were interviewed earlier in the process, to gain their view upon such a system's viability, value creation, and potential application in their daily learning environment. Similar to the first round of interviews, the discussions were individual, and based on a number of pre-defined questions. Before conducting the interviews, however, the students were presented to the idea, with a thorough explanation of the conceptual model, which was then used as a starting point for the discussion.

Based on the findings from the analysis, the project was discussed, and the research question was evaluated. The methodological approach is illustrated in figure 1 below.

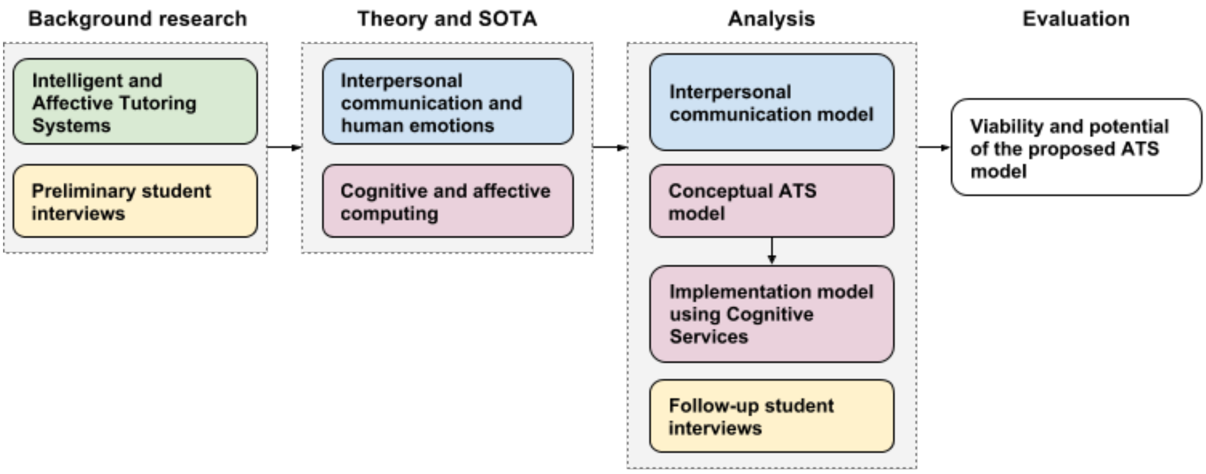


Figure 1: Illustration of the methodological approach.

3. Background research

To properly understand the different elements of the research question, additional background research will be made within a number of relevant areas, consisting of both primary and secondary research. This chapter will investigate intelligent and affective tutoring systems more in depth, before discussing the general attitudes and needs of students, in regards to their learning experiences and communication preferences.

3.1 Intelligent and affective tutoring systems

For decades, computers have been used to replace or support learning across a variety of educational situations. Two of the earliest so-called Computer-Assisted Instruction systems (CAI) are SCHOLAR (1970) and BIP (1976). SCHOLAR posed, answered, and evaluated learners' questions and answers, while BIP was used to assign programming tasks to students, based on their individual learning needs and competencies. Since then, learning technologies have developed into adaptive and interactive computer programs, which can provide advanced individualised tutoring (Ma et al., 2014). This individualisation is created by modelling the learners' individual traits, such as psychological states, learning characteristics, needs, and pace of learning. Today, these systems are known as *Intelligent Tutoring Systems* (ITS). ITS can adapt to the individual student, and guide them through different steps in a problem solving process, typically in form of feedback, hints, or error messages (Steenbergen-Hu & Cooper, 2014). Using ITS has numerous benefits over traditional classroom teaching, in the sense that they are always available, non-judgemental, and provide tailored feedback (Sarrafzadeh et al., 2008).

ITS incorporates numerous theories from different scientific fields, such as learning sciences, cognitive sciences, pedagogy, psychology, mathematics, linguistics, and artificial intelligence, and are designed to follow the practices of expert human tutors (Graesser et al., 2011). Throughout the years, ITS have been developed mainly within mathematically grounded academic subjects such as algebra, physics, economics and statistics, but have also been applied in e.g. nursing, medicine, law, programming, and military training (Irfan & Gudivada, 2016; Kulik & Fletcher, 2016; Ma et al., 2014; Steenbergen-Hu & Cooper, 2014).

ITS are considered to consist of four conceptual components, which must be implemented when designing such systems, as stated by Sottilare et al. (2013). First of all, the system must contain an interface which the learner can use to communicate, and through which the tutoring system can present and receive information. The other three components are three different models which must be incorporated in the backbone of the ITS; a domain model representing the knowledge that must be learned, a student model representing the student's knowledge, and a tutor model representing instructional strategies.

Ma et al. (2014) elaborates on the functionalities of ITS in relation to these instructional strategies and the different tutoring functions that the system can have: Presenting information to be learned, asking and/or answering questions, assigning tasks, providing feedback or hints, or in other ways provoking cognitive, motivational or metacognitive change. These functions must then be addressed and customised based on the student's psychological state, by taking into consideration e.g. knowledge, motivations, and emotions.

Benefits of using ITS

Many different researchers have documented the benefits that ITS can have on learning. Two of the most comprehensive analyses were conducted by Ma et al. (2014) and Steenbergen-Hu and Cooper (2014), which will be presented in this section.

Ma et al. (2014) conducted a review of 107 different ITS research papers, where they investigated the learning outcome of ITS compared to other traditional forms of learning; teacher-led large-group instruction, small-group instruction, non-ITS computer-based instruction, textbooks or workbooks, and individualised human tutoring. They found that using ITS was generally associated with greatly increased learning outcomes, compared to teacher-led large-group instructions, non-ITS computer-based instruction, and textbooks or workbooks. However, the ITS did not have a significant effect compared to neither individualised human tutoring nor small-group human tutoring. Regardless of how the ITS was used, i.e. whether it was used as a primary educational tool, supporting teacher-led instruction, supplementing after-class instruction, or aiding homework, it proved to have significant positive effects across all uses (Ma et al., 2014). These results were found across all levels of education.

A similar review was conducted by Steenbergen-Hu and Cooper (2014), however, their focus was exclusively on higher education. They analysed the use and effect of 22 different ITS on college students, and their findings corresponded well with those of Ma et al., (2014). Steenenbergen-Hu and Cooper also found that ITS overall had a positive on the students' learning. Furthermore, they found that ITS had an increased effect on learning in all application areas, besides human tutoring, as also concluded by Ma et al. (2014). Another aspect of Steenbergen-Hu and Cooper's research documented that the learning outcomes did not differ significantly for different subject domains, but that ITS did improve learning across all subjects. Based on previous research made, they argued that ITS have a larger effect on college students compared to lower levels of education.

Similar research was made by both Kulik and Fletcher (2015) and VanLehn (2011), again with similar findings. However, both made interesting additional findings, such as the robustness of ITS across time, place, and educational setting, and that the learning effects were reduced in poorly implemented ITS.

The analyses presented by Ma et al. (2014), Steenbergen-Hu and Cooper (2014), VanLehn (2011), and Kulik and Fletcher (2015) all provided uniform results. This is due to the adaptation and increased interactivity that ITS provides, compared to other forms of traditional classroom instruction and learning. More specifically, the greater immediacy of feedback, more response-specific feedback, greater cognitive engagement, increased opportunities for practice and feedback, more learner control, and individualised task selection, all contribute to ITS improved learning outcomes (Ma et al., 2014). Despite the generally proven increased effect of ITS, all analyses showed that human tutoring was the most effective learning method, thus highlighting the important roles of teachers and pedagogic, and the interpersonal relationship between teachers and students. According to Mao and Li (2010), human tutors' superiority to ITS, is due to the teachers' ability to identify and respond to the individual student's affective state. The key to enhancing the capabilities and effectiveness of ITS, it widely considered to be enabling the systems to adapt to the student's emotional state (Sarrafzadeh et al., 2008).

Affective Tutoring Systems

In recent years, the ITS have developed into systems capable of adapting to the cognitive and affective state of the students in the same way that human tutors do (Alexander, Sarrafzadeh & Hill, 2008). These types of ITS are known as Affective Tutoring Systems (ATS). Some researchers have referred to such systems as Emotion-Sensitive Intelligent Tutoring Systems (EITS), but from this point onward, this project will use the term ATS.

ATS can help avoid some of the negative emotional states in learning environments, which can have great impacts on the students' cognitive processes, such as information processing, communication processing, decision-making processes, and creative problem-solving processes. Subsequently, this will affect the effectiveness of their learning (Erez & Isen, 2002). Negative emotional states, such as anxiety, frustration, boredom, and loss of concentration, are also experienced in ITS, which can be a result of a mismatch between the individual student's character and learning needs, and the functionality and adaptability of the ITS (Malekzadeh, Mustafa & Lahsasna, 2015). Thus, it is important to help regulating the student's emotional state, and ensure increased engagement and a generally more positive attitude towards learning, thereby preventing or regulating negative states, to achieve a better learning outcome. By detecting emotional states, responding to these states, generating relevant tutoring strategies, and communicating with affect, this is exactly what ATS aim to do.

Today, sensors and input devices capable of emotion detection are more accessible and available than ever before, but according to Bahreini, Nadolski and Westera (2016), the recent developments in these devices are still underexploited. They state that such sensors offer great opportunities for many e-learning applications, including ATS. Besides the devices already being available to students, they add the possibility of gathering affective user data unobtrusively.

Cameras are considered to be to most underexploited sensor, and it is by many considered the channel which can extract the most extensive and accurate affect information, due to recognition of facial expressions (Bahreini, Nadolski & Westera, 2016).

According to Alexander, Sarrafzadeh and Hill (2008), there are two main methods which can be used to identify the affective state of users. The first method is detecting emotions based on physical effects, by using e.g. physiological signals, heart-rate, gestures, or facial expressions. The other method is concerned with what actually causes a person to feel a certain emotion, by taking into consideration factors such as personality, goals and previous interactions. The most optimal way of identifying students' emotions, however, is to combine these two methods (Alexander, Sarrafzadeh & Hill, 2008). Malekzadeh, Mustafa and Lahsasna (2015) further elaborate, and state that it is equally important to determine how to regulate the emotional state, and respond appropriately. The following section will describe how this is done in ATS.

ATS architectural model

Besides the four different components of traditional ITS (interface, domain model, student model, and tutor model), ATS system architectures also have an emotion information processing module, which is used in conjunction with the student model. Figure 2 illustrates the architecture of an ATS, based on Malekzadeh, Mustafa and Lahsasna (2015) and Sarrafzadeh (2008).

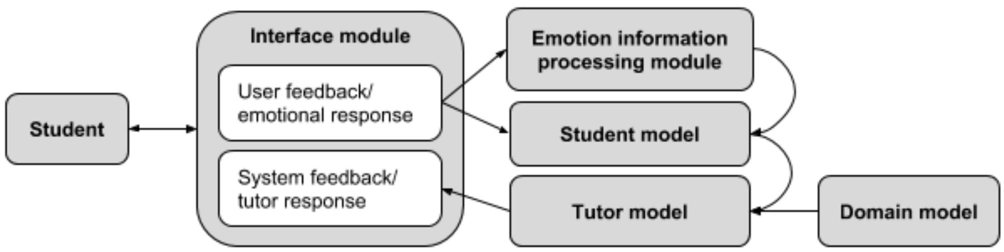


Figure 2: Architecture and components of an Affective Tutoring System.

The student interacts with the ATS's interface, which is where questions are asked by the student and responded to by the system, but it is also in this module where information about the emotional state is logged, before being sent to the emotion information processing module. The processing module then analyses this information and, based on the emotions detected from various modalities (e.g. physiological signals, facial expressions, and eye gaze), a weighing algorithm determines an overall emotional state. This state is often described in a valence-arousal model, which is a 2D-model used for emotion modelling. The student model also analyses the student's input in relation to the pedagogical state (information such as knowledge level and learning speed), and evaluates them according to the emotional state. Before constructing an appropriate response to the student, the tutoring module selects an appropriate tutoring strategy based on the student model. This strategy then determines the response, by drawing relevant content and material from the domain knowledge model. The feedback is then presented to the student in the interface. This feedback can be both domain dependent (related to course content) or domain independent (by showing empathy or encouragement).

3.2 Preliminary interviews

In this section, a group of university students will be interviewed in regards to their experiences and attitudes towards their educational environment, with a focus on the interaction, relationship, and communication between teacher and student, in both classroom teaching and project supervision.

Purpose

The purpose of the study is to gather primary data on both the students' educational needs and learning experiences throughout their studies, and to evaluate how this information correlates with the findings from section 2.1. More specifically, the interviews aim to elaborate on the following areas of interest:

1) *The general student experiences from class-room teaching, group work, and supervision.*
2) *The nature of the relationship between student and teacher.*
3) *The individual needs in regards to personal interaction and feedback, and to what extent these needs are fulfilled.*
4) *Technologies and methods currently used by the students to support learning.*

The aim is to gain further insights on the aspect of individual and personalised feedback, in relation to both interactions and interpersonal communication between teacher and students.

Interview guide

The interviews will be semi-structured, in the sense that the above mentioned areas will be discussed, based on a number of pre-defined open-ended questions. Four students from Aalborg University Copenhagen will be interviewed individually, and the interview process itself will be an open conversation. Based on the focus areas, the following questions are constructed and used as a starting point for the interview:

1) *What are your general opinions and experiences from class-room teaching and project supervision?*
2) *How would you describe your relationship with your teachers, and how does this relationship affect your learning experience?*
3) *How well do you feel like the teachers know you, and do they communicate appropriately to you as an individual?*
4) *How do you prefer to discuss and receive feedback in relation to reading material, lectures, tasks and assignments?*

5) *Are there sufficient resources for optimal supervision and feedback, or do you feel there is a need for additional tutoring?*
6) *How do you use technology to support the mandatory course content, project work, and supervision?*

The following sections will describe the findings from the conducted interviews, as well as highlighting some of the most noteworthy responses from the students.

Evaluation

This section will evaluate the interview responses in regards to the areas of interests mentioned above. Each aspect will be discussed in a general sense, and supplemented with relevant quotations, before presenting the key findings of the interviews.

General student experiences from class-room teaching, group work, and supervision

All four interviewees were studying at Aalborg University Copenhagen (AAU), and are thus working under the AAU model, which revolves around project-oriented and problem-based learning (POPBL). Their experiences with this model were mostly positive, as there was a general consensus about the increased benefits from project work. Especially the freedom to choose your own project focus increased the engagement and possibilities of specialisation in topics of interests, were highlighted by the students. Furthermore, they all responded positively about the university's focus on working in project groups, as it allows for constructive and inspirational discussions.

The nature of the relationship between student and teacher

There exists an open and friendly relationship between the students and their teachers, and the teachers are always approachable. Experiences from other educational programmes revealed that this is rather unique for AAU, and students from other universities and programmes do not have the same possibilities. However, as continuously mentioned by the interviewees, the students are not taking full advantage of this opportunity to get feedback from the teachers. One interviewee attributed this to the authority of the teachers, while another stated that it can also be a result of not wanting to expose one's lack of knowledge: *"If there is an expectation that you should have a certain skill set or knowledge, you do not always approach the teachers",* and elaborated that *"… it can be quite embarrassing to reveal your actual skill level".* Another interviewee confirmed this and said that *"When sitting in front of your professor… I cannot help thinking that it is some sort of presentation. You want to show that you can ask the right questions, and have the right answers."*

One of the reasons for the accessibility of help from the teachers, is the relatively small size of the programme in which the interviewees were enrolled. The number of students enrolled in a programme has a huge impact on different aspects of their relationship to the teacher, and subsequently the learning outcome. Based on the interviewees' experiences, it became clear that programmes and courses with 100+ enrolled students, had almost no personal communication with and feedback from the teachers. One student stated that questions were only answered extremely superficially, and that teachers responding to e-mails was a rare occurrence. In smaller groups, however, there is more time for individual presentations, discussion participation, and elaborate question answering. One interviewee described this as a *"quality-quantity trade-off".*

Another benefit of being a smaller group of students, is that the teacher can more easily adapt the content of the course to fit the semester projects, needs, and interests of the students. It also makes heavy course material more relatable, useful, and digestible. As a results, the interviewees felt more engaged in the class, and one student stated that he was more motivated to prove himself to the teacher, and that smaller groups *"forced concentration, focus, responsibility and participation".* Another student stated that *"the teachers are great at reading the students. This is because we are so few. You just connect in a different way.",* highlighting the more intimate relationship that exist between the teacher and smaller groups students.

The individual needs in regards to personal interaction and feedback

There was a general consensus among the interviewees, that their options for supervision and feedback was fulfilled. That being said, a number of interesting comments were made. First of all, was the importance of the group-oriented work. In their groups, the students have a much more informal relationship compared to within the class room with a teacher present. This allows for getting answers to some of the questions that they would otherwise hesitate to ask the teacher, as discussed in above. One of the interviewed students stated that group work and discussion also let you ask questions without sounding stupid or worrying about wasting the time of all other lecture attendants. He continued, and said that it also makes it easier for shy or nervous students to get answers to their questions. Another interviewee elaborated on this: *"… an essential part of it [group work, ed.] is to be able to spar with each other. You are more inclined to ask questions, and not afraid to sound stupid, which you also learn a lot from".* While smaller and less essential questions are discussed within project groups, the teachers are used more for general supervision.

As mentioned above, smaller classes are more considerate than larger classes, regarding adjusting the content to fit the individual students. One student, however, claimed that he still thinks that *"the individual needs, knowledge, and learning speed is not always considered, especially in regards to the educational backgrounds of the students."*

During class, the students generally benefit from feedback from both fellow students and teachers. However, there seemed to be slightly opposing views, as some mentioned feedback from fellow students as often being superficial and unconstructive, while others found the feedback beneficial and relevant.

In regards to personal assistance and feedback, different interesting issues rose from the interviews. It became evident that having one-on-one individual tutoring, was highly beneficial for the students, which they found very motivating and it *"had a huge impact in regards to improving and challenging yourself".* This interpersonal interaction with the teacher was rarely possible due to the number of students in class, compared to the time required of the teacher to provide sufficient individual feedback. One student gave an example hereof: "*In more practical classes, such as programming, you would benefit from having an additional teacher. We were 30 students, and our teacher was unable to sufficiently help all of us during class."*

Technologies and methods currently used by the students to support learning

In addition to the support offered by teachers/supervisors and fellow students, as in the two previous sections, the interviewees all used modern technology to assist their learning. Search engines were often used to find answers to simpler questions, while encyclopaedias and online databases were used for more scientific literature. One student specifically mentioned the usefulness of online educational videos, where experts elaborate on a certain topic. He argued that it sometimes helps to hear someone else rephrasing the course content, and that it can be beneficial to listen to experts explaining a specific topic in-depth, which was only sporadically mentioned in the course. Another student explained how the university establishes online discussion boards, and encourages students to use these if they have any questions, however, he stated that students never used this forum.

Key findings

Based on the above mentioned responses, the following list shows the key findings from the interview:

- High satisfaction with the POPBL model, and great benefits from working in project groups.
- Most university programmes do not have an open relationship to their professors, and their resources can be limited due to the large number of students enrolled in a course.
- Smaller classes enable more tailored content, which motivates student participation, and makes course material more relatable.
- There is a trade-off between the quality in personal feedback from teachers, and the number of students enrolled in a course.
- Interpersonal communication and feedback from the teacher had a great impact on student motivation and learning outcomes, especially in more practical classes.
- Individual needs, knowledge, and learning speed is not always considered within the classroom.

- Despite having the opportunity to always reach out to the teachers, the students feel certain inhibitions related to the expectations from the teacher, and have a fear of demonstrating a lack of knowledge.
- The informal relationship between students in project groups, allowed for discussion of issues and questions which would not be discussed with the teacher.
- Different technologies, such as search engines and scientific databases, are used by all the students as supplementary learning resources.

As became evident from the previously discussed academic research, that using ATS has a positive effect on the learning outcome regardless of the educational level. The research also showed that ITS and ATS are most effective when compared to larger classes, rather than small-group human tutoring or individual tutoring. The interviews conducted confirmed these observations, which, despite a general satisfaction with the educational environment, proved the need for additional tutoring in higher education.

Personal feedback is considered an extremely important aspect of learning, and while personal feedback and partially adapted course content is applied in smaller classes at AAU, the state within larger classes cannot be considered personal at all, and is far off the optimal learning environment. By implementing ATS, the barrier of fearing to ask questions and demonstrate a lack of knowledge, can be eliminated. This would help motivating students, increase their focus and concentration, assists their learning, and improve the overall effectiveness of their education. Furthermore, it would assist students who are shy, nervous, or asocial, and make them more integrated in the classroom. ATSs could help by taking into consideration the individual student's knowledge, learning speed, and learning needs, and provide questions and answers tailored to each individual, no matter the size of the class. However, it appears that ATS would be especially helpful in programmes and universities where group-work does not play as prominent a role as it does at AAU.

3.3 Chapter summary

The first topic addressed in this chapter was that of Intelligent Tutoring Systems (ITS), i.e. tutoring systems capable of adapting content and feedback, by modelling the individual learner's traits, such as psychological states, learning characteristics and needs. Research on ITS was then presented, demonstrating the benefits of using ITS in education. Until recently, ITS have been limited in regards to properly analyse and adapt to the deeper cognitive and emotional states of the learner. Today, tutoring systems have become more emotionally intelligent, and such systems are today known as Affective Tutoring Systems (ATS). It was briefly investigated how to obtain affect information, before describing an architectural model of ATS.

The second part of the chapter investigated the learning experiences and attitudes among university students. This was done by conducting four semi-structured interviews with students from Aalborg University Copenhagen. The students were asked about their general educational experiences, the nature of the relationship with their teachers, and their needs for personal interaction with and feedback from teachers. The responses showed a general satisfaction with the learning environment, however, all interviewees also mentioned problems within the educational system in regards to personalised content and feedback. The responses also revealed that the students sometimes feel certain inhibitions towards approaching their teachers. It became evident that technological solutions such as ATS, must be further investigated, as they have the potential to solve some of the existing barriers and problems within learning effectiveness in higher education.

4. Interpersonal Communication and Human Emotion

As previously mentioned, research has shown that the affective tutoring systems are inferior to small-group and individual human tutoring, due to the superiority of humans' adaptation to affect. To investigate what communication elements and approaches such an interaction incorporates, it is relevant to look at interpersonal communication.

This chapter will present a number of different aspects of interpersonal communication, starting with arguments for why communication is important. Then, key elements of the communication process are investigated, before presenting different theories on interpersonal communication, and how these elements are used within a number of communication models. The chapter also covers the role and importance of emotions in the communication process, to further investigate the understanding and communication of affect. Finally, based on extensive research, an interpersonal communication model will be constructed for later use in the analysis.

4.1 Why do we communicate?

Generally speaking, communication is the process of transmitting information and common understanding from one person to another (Keyton, 2011). This transfer of information and meaning can be either deliberate or accidental, but whenever a person does or says something and others observe and attribute meaning to the action, communication is taking place (Gamble and Gamble, 2011). Cheney (2011), however, states that if the exchange of information does not lead to a common understanding, there is no communication.

When discussing the domain of communication, Lane (2008) divides it into two different distinct types; intrapersonal and interpersonal communication (as depicted in figure 3). Intrapersonal communication is what takes place within an individual. Examples of this type of communication are internal dialog (a conversation within oneself) and self-talk (internal talk that is specifically about oneself). The other type, interpersonal communication, involves at least two people who establish a communicative relationship. The focus area of this project will be on interpersonal communication.

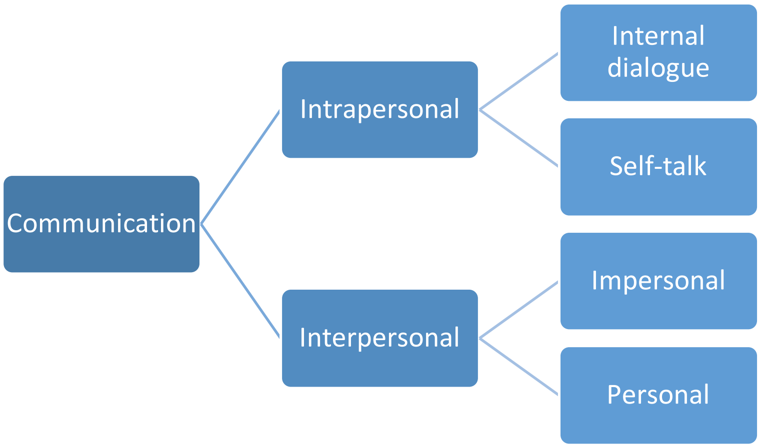


Figure 3: Different categories of communication (Lane, 2008).

Interpersonal communication is considered to be an ever-changing process which occurs when people establish and interact in a communicative association (Gamble and Gamble, 2011). Thus, any type of communication with two or more people involved is considered interpersonal. Bochner (1989) states that for communication to be considered interpersonal, three criteria must be fulfilled. At least two communicators must take part in the interaction; both must act as both subject and object; and their actions must embody each other's perspectives towards themselves and the other. He further elaborates that each communicator is considered to be *"both a knower and an object of knowledge, a tactician and a target of of another's tactics, an attributer and an object of attribution, a codifier and a code to be deciphered"* (Bochner, 1989, p. 336).

When engaged in interpersonal communication, people have the power to affect each other as interconnected partners in a relationship, in terms of both creating, maintaining, and dissolving relationships (Lane, 2008). The quality of an interpersonal relationship can be described along a continuum ranging from *impersonal communication* to *intimate communication* (Gamble and Gamble, 2011). Similarly, Lane (2008) also describes interpersonal communication as a continuum from *impersonal* to *personal communication*, even though it cannot be derived from figure 3 itself. Impersonal communication is considered to be stereotypical, i.e. when the receiver is treated as an "object" or imposed with a classic "role", while personal communication occurs when the interaction is based on the uniqueness of the other person. As a result, the more personal the interaction is, the more interpersonal the relationship is considered to be. When engaging in interpersonal communication, the goal is to treat and respond to each other as unique individuals with a distinct relational culture (Miell, 1984).

Rubin, Perse and Barbato (1988) identified six motives for interpersonal communication; pleasure (fun), affection (caring), inclusion (sharing feelings), escape (filling time to avoid other behaviours), relaxation (unwinding), and control (power). According to Lane (2008), if interpersonal communication fulfils these motives, we are likely to be satisfied with our interactions. However, how a person approaches and processes interpersonal communication is determined by context, culture, gender, the environment, roles, needs, history, and individual goals (Gamble and Gamble, 2011; Lane, 2008).

4.2 Principles of communication

It is generally understood that there are certain principles of communication which are important to consider when communicating with others, despite their cultural background. The following descriptions are based on the principles presented by Fujishin (2008) and Lane (2008). First of all, *communication is transactional*, which means that when participating in face-to-face conversation, the participants sends and receives messages simultaneously. Secondly, *communication is irreversible* in the sense that once a message is uttered it cannot be taken back or unsaid, and is thus always being interpreted by the receiver - a principle which is particularly applicable in computer-mediated communication, where a message can be permanently stored. Furthermore, *communication is an ongoing process* as it is difficult to distinguish between the beginning and end of communication, as it can begin as either an intrapersonal process, or based on a previous interaction with either the subject or someone/something else. The fourth principle is that *communication is constant* and it is impossible not to communicate, as behaviour is perceived and interpreted regardless of the sender's intention to communicate these signals. Whether a person is silent or speaking, laughing or crying, or expressing anger or joy, he is still communicating. In addition, *communication is learned* in the sense that communication patterns and behaviour are acquired and evolves over time. Such behaviour can be attributed to different factors such as our vocabulary, gestures, touching, appearance, and how we speak. Finally, *communication is creative.* Not only is it possible to communicate creatively with art, music and poetry, but communication can also be creative in relation to how, when, what and to whom we communicate.

4.3 The communication process

The communication exchange is often described as a process consisting of a number of different elements, but common for all are the sender and the receiver (Lunenburg, 2010). The **sender** is the initiator of the communication and the originator of the message. It can be a person writing an email, a public speaker, an advertising company, or any other creator of a given message that is to be transmitted. Before transmitting the message, it is encoded by the sender. **Encoding** is the process by which the sender composes the message using words, symbols, or gestures. The **message** is the idea, thought, or feeling, that the sender intends to communicate to the receiver, which can be communicated either verbally or nonverbally. The message is then transmitted via a **channel**, i.e. the medium by which it is communicated, which can be in any form interpretable by the human sensory system, such as face-to-face conversation, a phone call, an email, or a written letter. When the message arrives at its destination, it is decoded by the **receiver,** i.e. the recipient(s) of the message. The **decoding** is the mental process by which the receiver makes sense and extract meaning from the information received. When this initial transmission is complete and the message has been interpreted, the receiver responds to the message, a process known as **feedback**. The feedback can be verbal, nonverbal, or both, and it is via this process that the sender can estimate whether or not the message has been received and understood as intended. Feedback in communication is desirable, and only occur in two-way communication (Lunenburg, 2010).

Throughout the communication process, interfering forces, called **noise**, can distract from the communication and distort the message. It comes in a number of different forms, and can be either internal (psychological or physiological factors) or external (environmental factors). Examples hereof include interruptions, attitudes, and emotions. The quality of the communication is determined by the clarity in the different elements, and noise in any of the elements can reduce the effectiveness of communication (Keyton, 2011). The final element to consider in the communication process is the **context**, which refers to aspects such as the physical environment, time, cultural influences, gender, beliefs, values, and expectations. Communication is easier the more the participants' contexts overlap, whereas the interaction becomes gradually more difficult as the contexts diverge (Lane, 2008).

4.4 Communication models

To simplify and visualise the relations of the elements mentioned in the previous section, communication models are used to illustrate the communication process, as well as the interrelations between different components. Such models are useful in the sense that they enable analysis of how the individual components and their relations affect the outcome and process as a whole (Lane, 2008). Fujishin (2008), however, argues that while communication models are great visualisation tools, they are nowhere near comprehensive enough to describe everything about the communication process. The following sections will describe three different types of models; linear, interaction, and transactional models, and give examples of popular theories for each type.

Linear models

The simplest form of a communication model is linear of nature, meaning that the communication is unidirectional from the sender to the receiver, and thus omits the feedback or response from the receiver (Gamble and Gamble, 2013). An illustration of a linear model can be seen in figure 4 below.

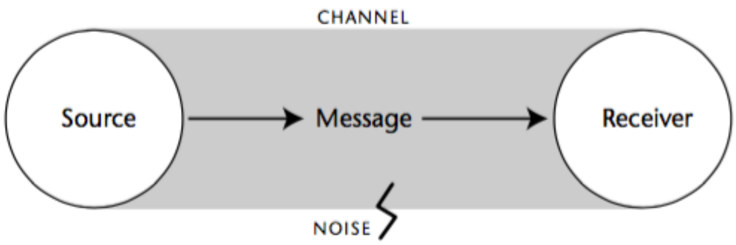


Figure 4: Illustration of a linear communication model, as per Fujishin (2008).

Widely considered one of the simplest ways of illustrating a communication process, is the Shannon-Weaver model of communication, as illustrated in figure 5 (Shannon and Weaver, 1949). The model was developed by Claude Shannon, electrical engineer at Bell Telecom Company, in 1948, and later popularised by Warren Weaver. The model was initially developed as a mathematical model to mirror the functionality of radios and telephones, but has since been adopted in a number of different domains as well, such as communication sciences, psychology, and social sciences. The Shannon-Weaver model was the first to introduce the concept of noise, representing the interference to the fidelity of the message, as experienced in a phone call or a radio transmission (Fujishin, 2008).

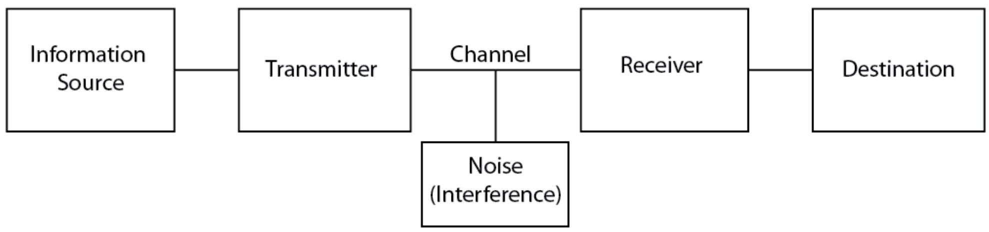


Figure 5: The Shannon-Weaver model of communication.

In 1960, David Berlo published his Sender-Message-Channel-Receiver (SMCR) model of communication, based on the Shannon-Weaver model (Berlo, 1960). Berlo introduced a number of different factors affecting the four main elements of the model, in relation to the efficiency of the communication. This model can be seen in figure 6 below.

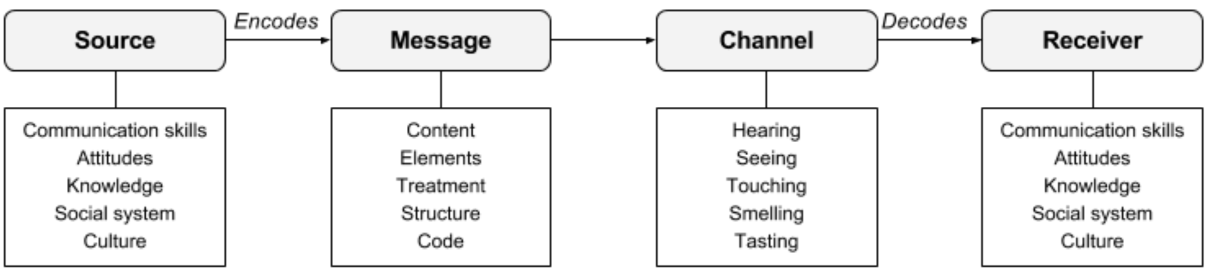


Figure 6: Berlo's SMCR model of communication.

While the simplicity of the linear model is a great advantage in visualising a basic message transmission, and contains several of the previously mentioned key elements, it is considered far to simple to describe the complexity of the actual processes. (Gamble and Gamble, 2013; Fujishin, 2008).

Interactional models

Following the criticism of the linear models, different bi-directional models were developed over the years, which became known as interactional models (Gamble and Gamble, 2013). As the name implies, interactional models visualise the interpersonal communication process as a back-and-forth process. Figure 7 illustrates the elements and interrelations present in an interactional model.

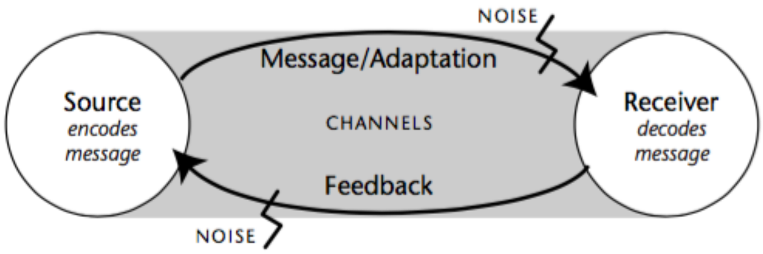


Figure 7: Illustration of an interactional model, as per Fujishin (2008).

An example of an interactional model, which breaks the classical one-way sender-receiver relationship, is the Schramm model. This model was proposed by Wilbur Schramm in 1954, and takes into consideration the interpersonal aspect, and illustrates the interaction as a circular communication process (Schramm, 1954). The Schramm model can be seen in figure 8.

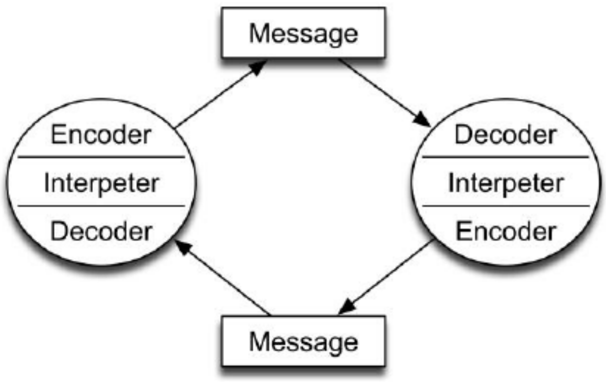


Figure 8: The Schramm model of interpersonal communication.

Besides the two-way communication, interactional models have the strength of taking into consideration the feedback element. When this is taken into account, the sender adapts the communication based on the feedback from the receiver, in order to properly communicate the message (Fujishin, 2008). Even though the interactional models are considered more accurate than linear models, it still does not quite capture the complexity and accurately reflect the processes of interpersonal communication (Gamble and Gamble, 2013).

Transactional models

While interactional models consider the communication process sequentially, transactional models see communication as a simultaneous process, as depicted in figure 9. Thus, the participants are referred to as *communicators*, rather than sender and receiver (Fujishin, 2008). In transactional models, the communicators are viewed as interdependent, unlike the previously mentioned models, where they are considered as independent elements in the system. Another important aspect of transactional models, is the influence of the individual backgrounds of the communicators, such as culture, psychological and physical factors, social factors, as well as demographic elements such as gender, and age (Lane, 2008).

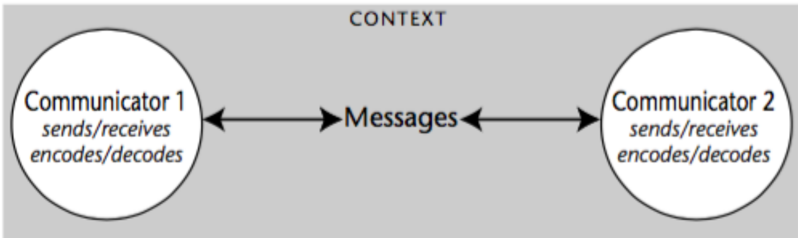


Figure 9: Illustration of a transactional model, as per Fujishin (2008).

By visualising communication as simultaneous, transactional models depicts the nature of a conversation in a much more realistic manner, compared to both linear and interactional models (Gamble and Gamble, 2013). Furthermore, the models contain all elements from the other two categories, while also including the context of the communication, thus representing the communication process in a much more holistic way, as the communicators continually influence each other (Fujishin, 2008). An example of a transactional model was presented by Stamp (1999), which will be further elaborated in the next section. A summary of the strengths and weaknesses of the three types of communication models can be seen in table 1 below.

Model	**Communication examples**	**Strengths**	**Weaknesses**
Linear	Telephone, television, radio, email.	Simplicity, applicability within electronic circuits.	Only considers one-way communication. No context or feedback.
Interactional	Instant messaging, individual speeches, class presentations.	Wider applicability.	Only considers one-way communication. No context.
Transactional	Face-to-face conversations, and any encounter in which meaning is co-created.	Most realistic depiction of interpersonal communication. Includes both context and feedback.	Does not apply to all forms of communication, e.g. broadcast, texting or posting.

Table 1: Strengths and weaknesses of linear, interactional, and transactional communication models.

Stamp's Grounded Theory Model of Interpersonal Communication

An interesting transactional model was presented by Glenn Stamp in 1999, which was based on grounded theory examining 288 research articles on interpersonal communication over the past 25 years (Stamp, 1999). This comprehensive and thorough model is based on seven main components, which were constructed on the basis of 17 sub-categories of interpersonal communication, as identified from the research. The components and their related categories can be seen in table 2, whereas the relations between the components are illustrated in figure 10.

The model depicts an ongoing transactional communication process between two people, as constructed by the seven different components. The first component, *culture*, is considered an overarching component, as it affects all aspects of the communication, by defining e.g. identities, social roles, norms, and the type of relationships developed. This cultural aspect is one of the drivers behind the second component, the *internal states*. The internal states are the interrelated aspects of perception, personality and cognition, which defines who we are, how we think, and how we make sense of the world (Stamp, 1999). The three processes of deception, compliance-gaining, and self-disclosure, composes the third component of the model, *interpersonal competencies*, which concerns interpersonal skills and competence levels. Besides these processes, the component is also fuelled by the internal states. The next component in the model is *communication apprehension*, which is based on the willingness or fear to communicate. Apprehension affects e.g. how we deceive or persuade another, and is considered to be a mediating component between the competencies and the message. The fifth component, *message behaviours*, represents the behaviours used to advance the previously mentioned components, i.e. the actual verbal and nonverbal messages sent in the communication process. The sixth component is the core category, that is the phenomenon which all the other categories are related to. In this model it is considered to be the *interaction/relationship* itself. During an interaction, conversations and conflicts takes place, and it is based on these actions that relationships are created, developed, and sustained. The final component of the model is the *interpersonal effect* that the interaction has on both communicators, and can thus be considered to be the feedback mechanism of the model. It encapsulates the reactions and emotions resulting from the interaction, which then affects the internal state of the participants, and subsequently influences the other components as well.

Component	Related categories	Description and examples of related variables/research topics
1. Culture	Culture	How culture affects communication, and how communication differs within various cultures; Acculturation, information acquisition, relationship types, relationship terms, social identity, intimacy, communication problems, communication constraints.
2. Internal states	Perception	The ways in which the communicator perceives the world; attitudes, attributional processes, person perception, interpersonal accounts, interpersonal expectations.
	Personality	Enduring personality characteristics of the communicator, and their effect on the communication process; Personality dominance, self-esteem, gender, assertiveness, anomia, communication style, individual differences, cognitive complexity, affective orientation, attachment style.
	Cognition	The internal state of the communicator; Information processing, pre-encoding aspects, pre-articulatory editing, memory, sequencing decisions, social cognition, motivation, knowledge, goals, cognitive editing, plans, intent.
3. Interpersonal competencies	Competence	Communication skills and competences; social perspective taking, role taking, social competence, development skill, social skill deficits, competence in different groups.
	Self-disclosure	Different facets of self-disclosure within interpersonal relationships; level, valence, and anticipation of self-disclosure, the relationship of trust, affection, and reciprocity to self-disclosure, disclosure styles, behaviours associated with disclosure.
	Deception	Interpersonal deception; deception as construct, detection of deception, deception in various relationship types, deceptive behaviours, relationship of deception to other personality characteristics.
	Compliance gaining	Strategies to gain compliance from another person, or the need to persuade within an interpersonal context; Compliance-gaining strategies in different types of relationships, resisting compliance-gaining strategies, persuasion, selecting compliance-gaining behaviours, effects of noncompliance, sequencing of strategies, verbal and nonverbal compliance-gaining strategies.
4. Communication apprehension	Apprehension	The construct communication apprehension; effects of communication apprehension, expectations of apprehensive others, non-verbal and verbal behaviours associated with apprehension, the relationship between communication apprehension and self-esteem.
5. Message behaviours	Message type	Different classes of messages that are primarily verbal in orientation; control, affection, empathy, ridicule, narratives, harmful speech, concern.
	Nonverbal communication	Aspects associated with non-speech related behaviour. Decoding nonverbal communication, body movement, space violations, illustrators, eye gaze, crowding, nonverbal encoding, motor mimicry, speech rate, touch, nonverbal immediacy behaviours.
6. Interaction/ relationship (Core category)	Initial interaction	Communication within initial interactions; initial interactional outcomes, attraction, homophily, dissimilarity, initial interaction scripts, language features during initial interactions, interactions with different types of persons.
	Conversation	The nature of conversation, conversational components, effects on conversations of particular devices; quantity of speaking time, conversational grammar, conversational turn taking, conversational structure, interruptions, pauses, interaction management.
	Conflict	Conflicts within interpersonal life; models of conflict, verbal conflict tactics, managing conflict, conflict strategies, mediation, argumentation, aggression, conflict in different contexts.
	Relationship development	Communication during the development of close relationships; social penetration process, relationship change, turning points, idiomatic communication in developing relationships, relationship escalation, impact of social networks on relationship development.
	Close relationships	Communication within close relationships; instrumental and expressive communication, patterns of communication, understanding, content themes, relationship satisfaction, play.
7. Interpersonal effect	Interpersonal effect	The interpersonal effects of messages, situations, or other variables; outcome of helping requests, anxiety-arousing messages, comforting messages, messages with multiple goals, guilt messages, stereotyping, consequences of speech styles.

Table 2: The seven components of Stamp's model and their 17 related categories (Stamp, 1999).

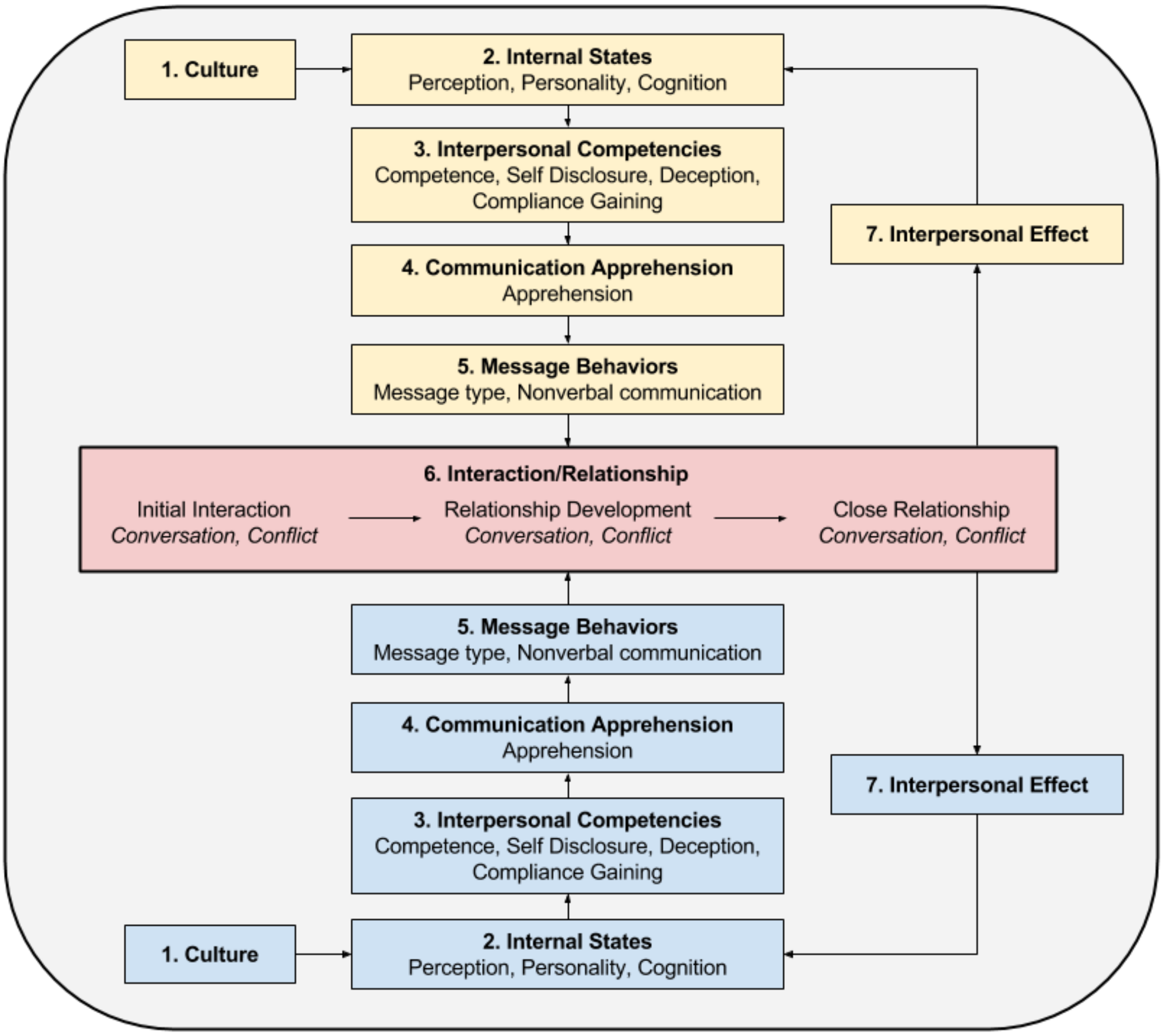


Figure 10: Stamp's grounded theory model of interpersonal communication.

4.5 Modes of communication

This section will present the two different modes of communication, verbal and nonverbal. Furthermore, the different types of communication signals will be discussed for both modes.

Verbal communication

Verbal communication refers to human interaction through the use of words or general messages in linguistic form, such as spoken and written communication (Chandler and Munday, 2011a). It concerns how we use words to create and convey meaning, which enables people to interact and allows them to *"define and classify; express our beliefs, attitudes, thoughts, and feelings; organize our perceptions; and talk about hypothetical events that occurred in the past, and events that may occur in the future"* (Lane, 2008, p. 126). This interaction allows for e.g. problem solving, relationship development, impression management, and fulfilment of basic human needs.

As per Ogden and Richards (1923), there exists an arbitrary relationship between words and the referents that they represent. Due to the nature of this relationship, a single word can be assigned different meanings depending on the context, but different words can also be used to describe the same entity (Fujishin, 2008). To avoid misunderstanding, language is said to conform to three types of communication rules; syntactic, semantic, and pragmatic rules. While syntactic rules define the arrangements of words, semantic rules refer to the meaning of the words. Pragmatic rules help interpreting verbal communication in a given context, and to determine the appropriate response within that context. The arbitrary relationship and the communication rules are critical components of verbal communication, as the meaning and ambiguity of words can greatly affect the (in)effectiveness of communication (Lane, 2008).

The aspect of *meaning* in verbal communication can be divided into two subcategories; the denotative and connotative meaning. The denotative meaning of a word relates to the very definition of the word, i.e. its "dictionary meaning", whereas the connotative meaning refers to the emotional or attitudinal response on an individual basis, as composed by feelings, images, and memories surrounding that word (Fujishin, 2008). Thus, denotation and connotation are extremely important to consider when communicating verbally.

Non-verbal communication

The other mode of communication, non-verbal, is defined as any form of communication other than verbal language (Chandler and Munday, 2011b). It should, however, not be considered secondary to verbal communication, as we communicate more through nonverbal communication than through verbal language. Nonverbal communication is widely considered to be the dominant communication mode, with some researchers demonstrating that between 65% and 93% of the meaning from messages is communicated with nonverbal cues (Ruggiero, 2000). This is related to the fact that the majority of nonverbal communication is transmitted unconsciously or unintentionally (Chandler and Munday, 2011b). Forni (2002) even argues that when used in conjunction with verbal communication, it is 12-13 times more powerful than the verbal message it accompanies.

While some aspects of the two modes are similar or related, certain characteristics distinguish the nonverbal dimension of communication from verbal. Fujishin (2008) argues that nonverbal communication is much more complex compared to its verbal counterpart, as it cannot be broken down and examined in discrete units (such as words and sentences). It is considered to be continuous and instantaneous, in the sense that e.g. surprises or disappointing messages immediately reflects physical and emotional response to information in e.g. face, posture, and breathing. Fujishin also states states that "*nonverbal communication conveys emotions and feelings much more effectively than words"* (Fujishin, 2008, p. 56). The aspect of conveying emotions was also denoted by the renowned American Psychologist, Michael Argyle, as he identified four main functions of nonverbal communication (Argyle, 1975):

1) Expressing emotion (affect displays).
2) Communicate interpersonal relationships and attitudes.
3) Support verbal interaction.
4) Presenting personality.

Compared to verbal communication, nonverbal communication is considered more universally understood, as it is still possible to derive some meaning from communication despite not understanding the spoken language. Certain specific nonverbal behaviours and expressions are considered to be universal, as per Paul Ekman (1978). However, the interpretation of nonverbal cues can also be culturally dependent, and communicated both intentionally and unintentionally, so the interpretation must be done carefully (Lane, 2008). The following sections presents the four different categories of nonverbal communication; kinesics, proxemics, haptics, and paralanguage.

Kinesics

The first category of nonverbal communication, kinesics, revolves around the study of body movement and positioning. It covers the areas of posture, gestures, facial expressions, and eye behaviour, and is an extremely important aspect of communication and interaction, due to the often multimodal aspect hereof.

While the posture refers to the positioning and orientation of the body, gestures are the visible acts used to communicate by the means of moving one or more body parts (Fujishin, 2008). The rather simple actions of standing and/or gesturing, can convey extremely strong messages, intentionally or unintentionally. While gestures are often considered to be used as a way of enhancing verbal communication, they can also be used as a primary method for conveying a message, e.g. when there is too much noise to use verbal communication (Kendon, 2007). Over the years, researchers have proven gestures to be a universal and natural form of expression, however, it is still considered to be governed by certain social conventions and cultural contexts (Fujishin, 2008; Kendon, 2007).

The other two areas within kinesics, facial expressions and eye behaviour, are also considered to be crucial components in our nonverbal behaviour. It is no coincidence that the vast majority of studies on human emotions have revolved around the visual cues given by facial expression and eye behaviour, since these are considered to be extremely potent sources of nonverbal communication, and functions as our prime communicators of emotion (Lane, 2008). Our internal emotional state is largely communicated through our facial expressions, and it is often done consistently and unintentionally (Fujishin, 2008). Facial expressions are constructed from the arrangement and activation of facial muscles, which is explained more detailed in section 4.6.

Proxemics

This category concerns the use of personal space and distance, and the study of proxemics can reveal many details about how we feel about ourselves and about others (Lane, 2008). Anthropologist Edward Hall defined four distances of personal space in Western culture, which can be used to imply the relationships between interacting people (Hall, 1966). The four spaces are:

1. Intimate space (0-0.45m): Reserved for intimate activities and conversation of confidential information, and only people with whom we share an intimate relationship are allowed.
2. Personal space (0.45m-1.2m): Used mainly for informal conversations
3. Social space (1.2m-3.6m): Appropriate for business discussions that are neither personal nor private.
4. Public space (3.6m-7.6m): Often used to communicate in class lectures, public speeches, or ceremonies.

The farther the distance is between two people, the more formal and impersonal their relationship and conversation. The exact distances may vary according to the setting, the people involved, the current emotional state of an individual, as well as a number of cultural background factors (Fujishin, 2008).

Haptics

Haptics refers to communication and interaction via the sense of touch, as is considered to be the most intimate form of nonverbal communication behaviour (Fujishin, 2008). The actual meaning conveyed with touch depends on a number of different factors such as the body part being touched, duration and method of the touch, the physical context, as well as the age, sex, and interpersonal relationship of the people involved. Research has shown that numerous emotions can be communicated solely with haptics, as exemplified by Hertenstein and Keltner (2006), who found that participants were able to decode anger, fear, distrust, love, gratitude, and sympathy via touch.

Paralanguage

The final area within nonverbal communication is that of paralanguage, sometimes also referred to as *vocalics*. It refers to *how* we say something, rather than *what* is being said (Lane, 2008). Paralanguage covers a number of different prosodic elements, such as vocal qualities and accents. These qualities can strongly affect how people perceive each other, and how an utterance is understood. It is possible to radically change the meaning of the spoken word by altering one's vocal qualities, such as the volume, pitch, speed, and intonation (Fujishin, 2008). A slowly speaking, lower pitched voice might seem calm and authoritative, whereas a person with high pitched voice who speaks at a higher rate might appear more excited. An incredible amount of information can be derived from paralanguage, as it is possible to identify or approximate individual factors such as ethnicity, age, as well as a number of personality characteristics, just by listening to how a person speaks. Vocal qualities also enable us to use and interpret speech as e.g. sarcasm or mockery, and allows for more sophisticated, sympathetic and emotional communication (Lane, 2008).

Even though the two modes of verbal and nonverbal communication are defined as separate entities, they are considered to be extremely closely related. Both Chandler and Munday (2011a), Fujishin (2008), and Lane (2008), argue that due to the ambiguous nature of nonverbal communication, it is preferred to interpret the two in conjunction, and should be seen not as distinct but as complimentary modes.

4.6 Human emotions in interpersonal communication

If interpersonal communication is to be emulated in an ATS, the emotional signals in the communication process is of course of highest importance. This section will discuss the relevance of understanding and using emotions in a communication context.

Definition and relevance of emotion

When discussing interpersonal communication, emotions play an essential part due to the reciprocal relationship of emotion and communication. Emotions are decisive for relationships building, while social relationships are the most important source for human emotions, as people are spurred to emotion from interpersonal interactions (Langlotz & Locher, 2013; Qadar, 2016). Many different definitions of emotion exist, but according to Lane *“emotions are feelings we experience that result from the interaction of physiology, cognitions, and social experience and that they significantly affect how we communicate with others and interpret others’ communication”* (Lane, 2008, p. 98).

The link between emotional communication and interpersonal relationship is elaborated upon by Andersen and Guerrero (1997), who lists six applying principles that forms this link;

1. Socially adaptive emotional communication is positively selected in the evolutionary process.
2. Socialization processes guide how individuals manage their communication of emotion.
3. Interpersonal schemata [scripts of ‘normal behaviour’], including goals, needs, desires, and expectations affect how and when emotion is experienced and communicated.
4. Interpersonal communication is the primary elicitor of most emotions.
5. An essential feature of the emotional experience is expression via interpersonal communication.
6. Emotions generate other emotions in interaction chains.

Langlotz and Locher (2013) divides these principles into two categories; The first three are related to the basic biological, socio-cultural, and cognitive connections; while the last three are related to the actual role that emotions play in interpersonal communication.

Expressing emotions is also a way of covering some of the basic human needs, e.g. expressing sadness resulting in social support and promoting personal bonds and intimacy (Mongrain & Vettese, 2003). Feelings and emotions affects when, how, and why we communicate, and emotion is thus considered to be a key element in interpersonal relations (Lane, 2008). Research has even shown that hardly any communication takes place without emotional involvement, and some even suggest that it is impossible to fully understand communication without taking this aspect into consideration (Altrov, 2013). Furthermore, Qadar (2016) states that using, managing and communicating emotions is an essential skill of human development. This is backed by Langlotz and Locher, who state that *"human sociality is fundamentally grounded in our ability to empathize and emote with others"* (Langlotz & Locher, 2013, p. 92). Another key aspect of communicating and experiencing emotions lies in our individual personalities, as three of the Big Five personality traits strongly influences our emotions (extroversion, neuroticism, and agreeableness; Lane, 2008).

Emotional signals and how they are expressed can vary greatly depending on the person, the culture, and the specific context in which the signals are sent (Hoey, 2017). As an example, heart rate might increase when a person is feeling angry, but the same response could be seen when a person is performing a physical activity. It can be extremely difficult to distinguish between these two behaviours, thus making the task of emotion recognition extremely difficult. According to Hoey (2017), emotion signals can be divided into two distinct categories; signals that are apparent to others, and those less or not apparent to others. The former includes signals such as facial expressions, gestures, body movement, and voice intonation. Less apparent signals are more physiological of nature and include respiration, heart rate, blood pressure, and electro dermal response.

Perception of emotion

A critical aspect to consider when discussing the use of emotions in communication, is how we each individually perceive each other and the world. By taking into consideration the individual differences of this perception, the general communication effectiveness can be improved.

Perception is the process by which we select, organise, and interpret sensory impressions in order to give meaning to the environment and interactions in which we participate (Pipas & Jaradat, 2014). The meaning extracted from these sensory inputs can then evoke certain emotions and even change the emotional state of a person (Maiocchi, 2015). As perception is essentially a subjective process, we all perceive the same stimuli differently, and thus communicate in a manner that facilitates our own understanding of the perceived.

The Perception Process

According to Lane (2008), the perception process has three distinct stages that occur almost simultaneously: selection, organisation, and, interpretation. The first stage, *selection*, relates to how we select from the surrounding stimuli of a given scenario. To sort the information, all stimuli are evaluated according to their salience (i.e. the interest, use, and meaning) and vividness (i.e. stimuli that are more noticeable). After selecting which information to process, the stimuli are organised on the basis of e.g. schemas, closure, and proximity and similarity. The final stage in the perception process is the interpretation stage, where we make meaning of the organised stimuli. This interpretation is influenced by both expectancy and familiarity, i.e. what we expect to perceive, and how familiar we are with the stimuli.

Pipas and Jaradat (2014) state that the perception is affected by a number of different factors in both the one perceiving (the observer), the object being perceived (the target), and the context in which the perception is made (the situation). Each observer has different attitudes, motives, interests, experiences, and expectations, which all contribute to the perception of an object or a situation (Langton & Robbins, 2006). Such individual differences in the characteristics of the observer can distort perceptions, and thus lead to misinterpretations of other's behaviour. Furthermore, the cultural background of the observer also plays a significant role in the perception process. The second factor affecting the perception is the characteristics of the target, such as loudness, novelty, movement, sound, size, or other attributes. In addition, nearby and surrounding objects will also affect how each individual object is perceived, as they are never viewed as isolated objects but always within the situation in which they are perceived. The situational aspect also concerns lighting conditions, the moment in time, location, temperature, noise, and other situational factors.

Perceptual errors

Each of the three mentioned stages are influenced by personal bias, which can lead to perceptual errors, such as selective perception, selective attention, or confusing fact with inference (Lane, 2008). These perceptual errors are due to the influence of human physiology, social roles and statuses, and individual cultural background (Pipas and Jaradat, 2014). Our perceptions are heavily influenced by physiology in regards to the acuity of our senses, as differing sensory capacity also means differences in perception. Furthermore, the biological rhythm of the body affects many different processes, such as body temperature, attention level, mood, and hormone levels. As a result, energy and stress levels also tend to differ. Another influential factor is that of social roles and statuses. Depending on the role and status of a person, a certain type of behaviour is expected in certain social situations. As the same person can play multiple social roles, it also influences the perception of that person. The third and most influential factor is considered to be the cultural background of the observer, as the values, norms, beliefs, symbols, and customs of a person, is a reflection of the organisational culture and subculture which he identifies with. It influences both the behaviour and attitudes of the person in question, but also how that person perceives and interprets others' behaviour.

Emotional cues

A number of different emotional cues and communication channels exist, i.e. emotions are communicated by a range of verbal and non-verbal cues. By being able to understand and decode these emotional cues, interpersonal communication becomes more efficient, which subsequently leads to a greater satisfaction and more willingness to communicate (Qadar, 2016). Planalp (1998) presented five different classes of emotional cues, very similar to the cues presented in section "Modes of communication". An overview of Planalp's classes can be seen in table 3 below.

Class	**Forms of realization**
Verbal cues	Language-specific emotion vocabularies. Metaphors. Speech acts. Emotional discourse practices, e.g. therapeutic discourse.
Vocal cues	Voice quality: low, loud, slow, fast, trembling, high-pitched, monotonous, animate voice.
Body cues	Animated, energetic movement. Physical actions: throwing, threatening, kissing, caressing. Gait: walking heavily, lightly, arm swing, stride. Body posture: stiff/rigid, droopy, upright. Gestures: emblems, clenching hands or fists.
Physiological cues	Blushing, pupil dilation, heart rate, breathing, skin temperature.
Facial cues	Facial expressions through forehead and eyebrows, eyes and eyelids, mouth and lips.

Table 3: Emotional cues, as presented by Planalp (1998).

A barrier to correctly interpreting emotional cues, is that of intercultural differences, as certain culturally display rules determine how and when to display emotions, and comes in the form of both encoding and decoding rules. By adhering to display rules, it is possible to e.g. simulate, intensify, de-intensify, or mask the actual emotion, and thereby communicate appropriately in the cultural context (Langlotz & Locher, 2013). On the other hand, violating such rules can lead to miscommunication, hostility, or even communication breakdown (Gallois, 1993). Tannen (1986) even states that similar rules and misunderstanding hereof also exist in intracultural communication, between people from different classes, regions, or gender. To avoid such misunderstandings, the expression of emotions is adapted based on the context and surroundings, however, due to the sometimes ambiguous nature of emotions, different people may react differently to the same displays (Langlotz & Locher, 2013).

Despite these differences, some considers human emotionality universal due to the measurability in biological processes, as certain stimuli can invoke e.g. increased heart rate, blood pressure, etc., which corresponds to a primary emotion (Langlotz & Locher, 2013). Additional factors in how emotions are experienced and communicated are individual context, physiology, and cognitions

(Lane, 2008). Barriers to communicate emotions can also be based on conscious or unconscious evaluation of the context in which emotions are communicated; the motive can be considered unethical, uncertainty of how the receiver will interpret or react to the message, the physical environment in which communicate takes place, inappropriateness due to social roles, or for reasons of privacy (Lane, 2008).

The importance of multimodality

As mentioned in the previous sections, communication signals are sent in different channels, and due to the complexity of the signals, they are preferred to be communicated multimodally, i.e. through more than one sensory channel. The phenomenon of multimodal communication is important in many different disciplines, including psychology, perception, and cognition, and thus also when studying the contribution and importance of emotions in interpersonal communication (Partan & Marler, 2005).

The multimodal aspect of emotional expression is especially important in interpersonal communication, since verbal communication is almost always accompanied by appropriate gestures and facial expressions, especially in case of spontaneously occurring emotional expression (Sherer & Elgring, 2007). According to Bänziger, Grandjean and Scherer (2009), a number of prototypical sets of expression configurations can be defined, which are used to not only understand signals sent by familiar persons, but also used to understand the expressions of complete strangers with little contextual information available. They consider these prototypes as core skills used in interpersonal communication, as observed expression patterns are used to evaluate e.g. the authenticity of the messages received. Discrepancies between different channel signals is often seen as a sign of deception from the sender. To better understand interpersonal communication, relationships, and emotional expression, it is thus necessary to view the interaction from a multimodal point of view, and thereby taking into consideration all elements of the complex emotional signals (Langlotz & Locher, 2013). The core skill related to understanding the overall emotional meaning of a message, is considered to be the ability to resolve discrepancies in messages, that is to evaluate and weigh the individual channels in a multimodal message (Gallois, 1993).

Analysing emotion expression unimodally, on the other hand, is considered to be highly complicated, and often results in misunderstandings, miscommunication, and misleading test results. This could also be a result of modulation or withholding of information in a given channel, which could have been interpreted more accurately with information from additional channels. However, research has shown that some channels (especially facial expressions) are more easily decoded and understood unimodally than others. Partan and Marler (2005) suggest that multimodal signals should be studied by taking both the individual channels and the composite signal into consideration.

Throughout the years, a number of studies have been made on multimodal emotion recognition. Some of the most important tests develop include DANVA (Nowicki & Duke, 1994), MERT (Bänziger, Grandjean & Scherer, 2009), ERI (Scherer & Scherer, 2011) and perhaps most notably the PONS test (Rosenthal et al., 1979). While all of these tests include both visual and auditory cues, the exact application hereof differs. The results of the tests, however, all prove that multimodal emotional expressions are more accurately understood than their unimodal counterparts.

4.7 Theories on human emotion

Throughout the years, researchers from different academic fields have tried to conceptualise human emotions. This section will present some of the most renowned theories, such as Paul Ekman's basic emotions, and Russell's circumplex model.

Ekman's basic emotions

Many different theories on human emotions exists, but one of the pioneers within the field is American psychologist Paul Ekman. Based on intercultural studies, Ekman found six basic emotions that were universal; *anger, disgust, fear, happiness, sadness,* and *surprise* (Ekman, 1978). He states that each of these emotions can be expressed using facial expressions, and are universally understood. In 1978, he presented the *Facial Action Coding System* (FACS), which is a widely used measurement system for facial expressions (Ekman, 1978). The system consists of numerous *Action Units* (AUs), which are based on the movement of the human facial muscles. While FACS itself does not indicate emotional states, it can be used as a basis for further analysis using FACS codes from which emotions can be inferred. FACS was updated in 2002 to accommodate head movements and eye positions, while also refining some of the AU combinations (Ekman, Friesen and Hager, 2002). This research added to the list a number of additional basic emotions, however, these were unrelated to the facial expressions; *amusement, contempt, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure,* and *shame*. In recent years, FACS have been used in both neuroscience, computer vision, computer graphics and animation (Poria et al., 2017).

Russell's Circumplex Model of Emotions

A number of dimensional models of emotions have been proposed throughout the years, and have since been widely adopted. One such dimensional theory is Russell's *Circumplex Model of Emotions*, which distributes 28 emotions in a two-dimensional space consisting of arousal (vertical) and valence (horizontal) (Russell, 1980). An illustration of the model can be seen in figure 11 below. This model provided a universal method of understanding and analysing affect, to great interest in e.g. psychology and learning. Within a learning context, the preferred emotional states are those found in the first quadrant, i.e. those pleasant with high arousal. In other words, these emotional states have the highest positive activation, and thus the highest learning outcome.

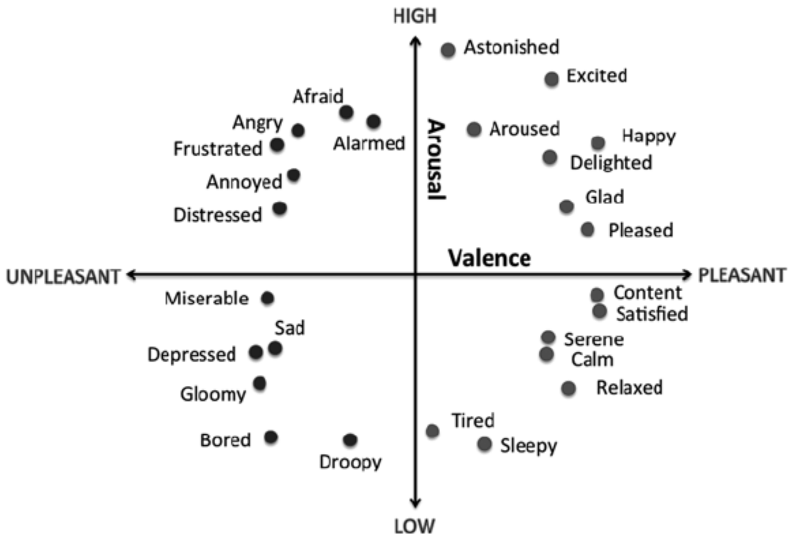


Figure 11: Circumplex model of 28 emotions, according to valence and arousal. Illustration by Nesbitt et al. (2015), based on Russell (1980).

Plutchik's Wheel of Emotions

Another widely renowned dimensional model is the *Wheel of Emotions*, as introduced by Robert Plutchik in 1980 (Plutchik, 1980). The model is based on eight primary bipolar emotions; *joy-sadness, anger-fear, trust-disgust,* and *surprise-anticipation*. Based on the intensity of each of these basic primary emotions, more advanced emotions arise such as *annoyance, terror, boredom,* and *ecstasy*. Furthermore, Plutchik also theorised a number of primary, secondary and tertiary feelings composed of two emotions. These constructs were denoted *dyads.* The vertical dimension in the model represents intensity, while the radial dimension represents similarities among the adjacent emotions. An illustration of the complete model can be seen in figure 12.

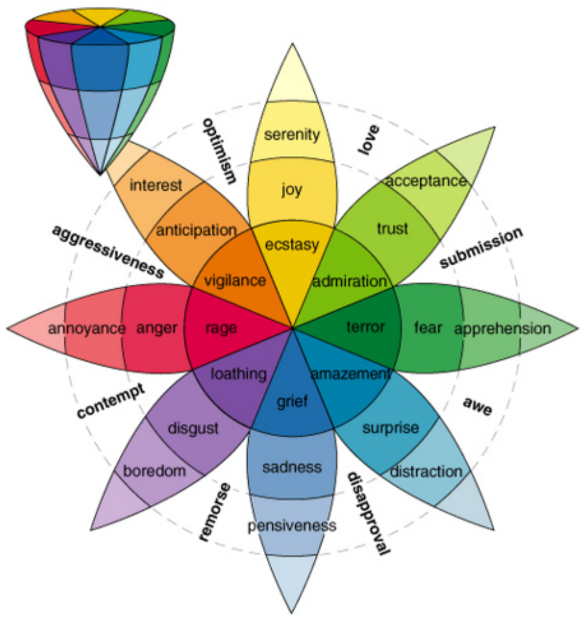


Figure 12: Plutchik's Wheel of Emotions (Plutchik, 1980). The color saturation indicates the intensity of the primary emotions, and primary feelings are denoted between the branches.

Parrot's classification of emotions

Another, even more elaborate theory was presented by Parrot (2001). Like Plutchik, Parrot describes emotions in three dimensions, with the six primary emotions being *liking, joy, surprise, anger, sadness,* and *fear*. Each primary emotion then encompasses a number of secondary emotions, which then again are broken down into tertiary emotions. The tree-structure of the theory, and the full list of identified emotions can be seen in table 4.

Primary emotion	**Secondary emotion**	**Tertiary emotions**
Love	Affection	*Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality*
	Lust	*Arousal, desire, lust, passion, infatuation*
	Longing	*Longing*
Joy	Cheerfulness	*Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria*
	Zest	*Enthusiasm, zeal, zest, excitement, thrill, exhilaration*
	Contentment	*Contentment, pleasure*
	Pride	*Pride, triumph*
	Optimism	*Eagerness, hope, optimism*
	Enthrallment	*Enthrallment, rapture*
	Relief	*Relief*
Surprise	Surprise	*Amazement, surprise, astonishment*
Anger	Irritation	*Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness*
	Exasperation	*Exasperation, frustration*
	Rage	*Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment*
	Disgust	*Disgust, revulsion, contempt*
	Envy	*Envy, jealousy*
	Torment	*Torment*
Sadness	Suffering	*Agony, suffering, hurt, anguish*
	Sadness	*Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy*
	Disappointment	*Dismay, disappointment, displeasure*
	Shame	*Guilt, shame, regret, remorse*
	Neglect	*Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult*
	Sympathy	*Pity, sympathy*
Fear	Horror	*Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification*
	Nervousness	*Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread*

Table 4: Parrot's classification of emotions (Parrot, 2001)

Another approach to modelling human emotion was taken by Hugo Lövheim, as presented in his *Cube of Emotion* (Lövheim, 2012). Lövheim defines eight basic emotions in a three-dimensional space, according to their relation to the signal substances dopamine, noradrenaline and serotonin, also known as the monoamine neurotransmitters. As an example, joy is constituted by high values of dopamine and serotonin, but with low values of noradrenaline, as illustrated in figure 13.

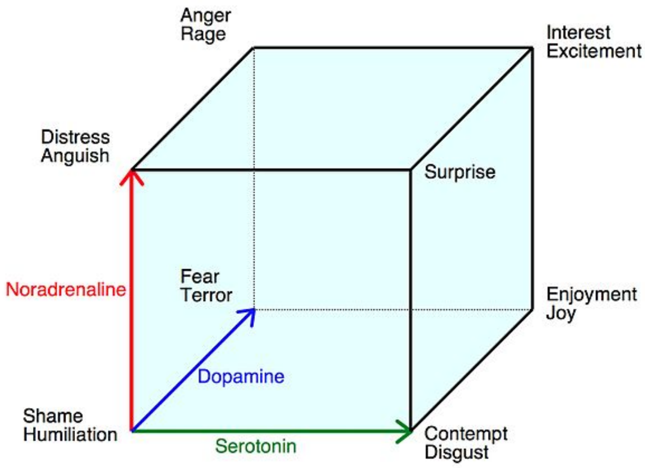


Figure 13: Lövheim's Cube of Emotion (2012).

An overview of the emotion theories presented throughout this section, can be seen in table 5.

Theory	Primary emotions
Ekman (1978). Basic emotions.	Anger, disgust, fear, happiness, sadness, surprise.
Russel (1980). Circumplex model of emotions.	Exited, astonished, delighted, happy, pleased, content, serene, calm, tired, bored, depressed, miserable, frustrated, annoyed, angry, afraid, alarmed.
Plutchik (1980). Wheel of emotions.	Joy, sadness, anger, fear, trust, disgust, surprise, anticipation
Parrot (2001). Emotions in Social Psychology.	Liking, joy, surprise, anger, sadness, fear
Lövheim (2012). Cube of Emotion.	Shame/humiliation, Distress/anguish, Fear/terror, Anger/rage, Contempt/disgust, Surprise/startle, Enjoyment/joy, Interest/excitement.

Table 5: Overview of renowned theories of human emotion.

4.8 Interpersonal Communication Model

Based on the research presented throughout this chapter, a model illustrating the different elements of interpersonal communication is created. The model is based on the overall structure of Stamp's Grounded Theory Model of Interpersonal Communication, but includes elements from many of the other research areas investigated as well. The model is illustrated in figure 14 below.

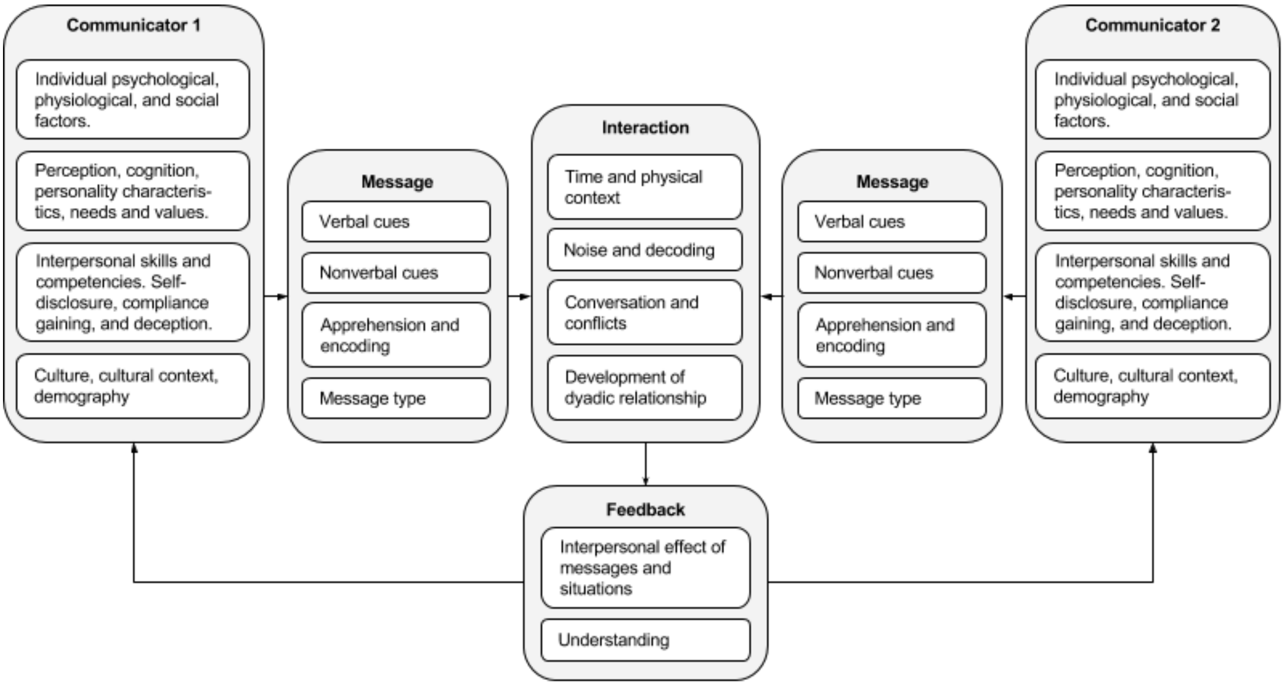


Figure 14: Summarised model of interpersonal communication.

This model will later be used to discuss the use of cognitive and affective computing within affective tutoring systems, in regards to how it can emulate human personal tutoring.

4.9 Chapter summary

After presenting a number of principles of communication, the general communication process was investigated, with a focus on the key elements which this process is considered to consist of; sender, encoding, message, channel, decoding, receiver, feedback, context, and noise. This was followed by a description of different communication models incorporating these elements; linear, interactional, and transactional communication models, including examples of each type of model. From the research, it became obvious that each of the mentioned components of the communication process must be taken into consideration when analysing interpersonal communication, in order to see the process holistically.

Other important aspects of interpersonal communication are the two modes of communication; verbal and non-verbal. When communicating a message, it is encoded and sent with a number of different cues. Examples of such cues are words and linguistics (verbal), gestures, posture, facial expressions, proxemics, and haptics (non-verbal). As each mode has both advantages and disadvantages over the other, interpersonal communication often makes use of both modes simultaneously. Often, a message is sent with multimodal cues, which increases the effectiveness of communication.

Then, the influence emotion has on interpersonal communication was investigated, as it significantly affects when, how, and why we communicate, and how we interpret others' communication. It is widely proven that no communication takes place without some emotional involvement, and the two areas are considered to be linked through basic biological, socio-cultural, and cognitive connections. The expression of emotions is based on the context of which the communication takes place, and how and when the emotions are expressed are modulated by different display rules. On the other hand, emotion-related stimuli are perceived differently from person to person, as the perception process is subjective, based on each individual's selection, organisation, and interpretation of the stimuli. This process is also affected by physiological and biological factors, as well as personal bias, e.g. such as selective perception, selective attention, and confusing fact with inference, which all lead to perceptual errors. Research also highlighted the importance of multimodality when communicating with affect, due to the complexity of emotional signals. In addition, a number of different theories on human emotion were investigated, as presented by Ekman (1978), Russel (1980), Plutchik (1980), Parrot (2001), and Lövheim (2012).

Finally, a model was constructed, describing the interpersonal communication process, by combining all information from the presented research. This model will be used for further analysis in chapter 6.

5. Cognitive and Affective Computing

The previous chapter highlighted the complexity and capacity of the human brain in regards to communication and emotion processing. To properly emulate interpersonal communication, it is thus relevant to look at how the neural processes can be recreated in computers. This chapter will present an elaborate investigation of Cognitive and Affective Computing. It will be discussed how the technologies are applied, in regards to their underlying architecture, and the functionality of the systems. Furthermore, the current state of the art of both cognitive and affective computing will be evaluated.

5.1 A new era of computing

According to IBM Senior Vice President, John E. Kelly, the evolution of information technology is approaching a new and third era, that is radically different from the previous eras (Kelly & Hamm, 2013). The first era, *The Tabulating Era*, was all about mechanical tabulating machines, used to organise data and perform calculations and counting based on this data. Tabulating machines were used in a number of different application areas, however, predominately in accounting. The Tabulating Era began in the nineteenth century, and continued into the 1940's, where information technology entered its second era: *The Programmable Era*. Focus now shifted towards storing and processing information digitally, and programming the software to execute logical sequences of steps, resulting in much more powerful systems (Kelly & Hamm, 2013). In its early years, these systems enabled computation processes such as message encryption and decryption, while more advanced systems have played a key part in space exploration in the second half of the century. Today, we are on the brink of the third era of computing: *The Cognitive Era*. Kelly states that cognitive systems will evolve from their interactions with both data and humans over time, from which they will be able to learn and improve their capabilities and accuracy, thereby extending what either humans or machines could do on their own (IBM, 2017a; Kelly & Hamm, 2013). The relationships between man and machine will blur, and the interactions will augment our senses and lives in ways that were previously unimaginable (Roe, 2014).

5.2 What is cognitive computing?

Before diving more into the area of cognitive computing, it is important to take a step back and understand the concept of cognition itself. Cognition is a term describing different conscious mental activities and processes, such as the activities of thinking, understanding, learning, and remembering, as experienced through thought, experience and the senses (Merriam-Webster, 2017; Oxford Dictionaries, 2017). Thus, it encompasses the core psychological processes of human thought and realisation.

Due to technological advancements in recent years, machines are now able to understand information, learn, reason, and act upon that information. The computational power and algorithms have reached such a level of sophistication, that the computers now appear to have the capability of actually *thinking* (Cognitive Computing Forum, 2014). This has led to the notion of *cognitive computing*. Many different definitions of cognitive computing exist, but one of the most concise descriptions has been made by Rouse (2016), who states that it is the simulation of human thought processes in a computerised model.

Mimicking human cognition in a computer, however, is easier said than done. Already in stage of representing and storing information, the computational process bears little or no resemblance to the corresponding neurological processes of the human brain (Raghavan et al., 2016). Earley even argues that the term is a misnomer, since cognition *"describes the act of thinking, and computers neither think nor understand in the literal sense"* (Earley, 2015, p. 12). In cognitive computing, the mind is considered a highly parallel information processor, in which various models, algorithms, and technologies are used to transform and reason upon information. Unlike traditional computing systems, cognitive systems are adaptive, learn and evolve over time, and even takes contextual factors and their surrounding environment into consideration, when computing and acting upon information (Raghavan et al., 2016). Enabling technologies include data mining, pattern recognition, machine learning, computer vision, natural language processing, robotics, big data, and cloud computing (Cognitive Computing Forum, 2014; Raghavan et al., 2016; Rouse, 2016). Systems utilising machine learning can continually acquire knowledge from new data, and are even able to deal with incomplete, unstructured and ambiguous information, thus eliminating the previous *brute force* approaches to artificial intelligence (Raghavan et al., 2016; Rouse, 2016). Possibly the most widely know application of cognitive technology today is IBM's *Watson*, who confidently beat two previous winners of the game show *Jeopardy*, a technology which is further elaborated in section 5.6.

5.3 Deep learning and neural networks

One of the most significant reasons behind the progress of cognitive and affective computing, is the re-emergence of deep learning. Deep learning is the foundation of applications across many different cognitive computing subdomains, such as image, video, speech, and text analysis (Schmidhuber, 2014). This section will present the concept of deep learning, with the main focus on the most widely applied deep learning method today, neural networks.

Deep learning

According to Guo et al. (2016), deep learning is *"a subfield of machine learning, which attempts to learn high-level abstractions in data, by utilising hierarchical architectures"* (Guo et al., 2016, p. 27). Similarly, Deng and Yu (2014) defines it as *"A class of machine learning techniques that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification"* (Deng and Yu, 2014, p. 199-200). Throughout the years, leading deep learning researchers have demonstrated the success of deep learning in both computer vision, phonetic recognition, voice search, hand-writing recognition, robotics, speech recognition, and many other domains (Deng and Yu, 2014). Some of the key factors for the emergence of deep learning in recent years, have been the increased processing power in e.g. GPU's, a significant decrease of computing hardware prices, the increased availability of training data, and the advances in machine learning algorithms (Goodfellow, Bengio & Courville, 2016; Guo et al., 2016).

In its early days, the domain of deep learning was intended to be computational models of how the brain learns. Today, the neural aspect of machine learning still plays a large part in engineering deep learning systems, however, some deep learning researchers are not concerned with neuroscience at all, but mainly focuses on fields such as information theory, linear algebra, and probability (Goodfellow, Bengio & Courville, 2016). As a result, today the notion of deep learning appeals to a more general principle of learning multiple levels of composition.

The notion of *deep* learning relates to the biological processes of the human brain which is organised in a so-called deep architecture. This means that stimuli or inputs are represented at multiple levels of abstractions, where each level corresponds to a different area of the brain's cortex (Bengio, 2009). In addition, information is processed in a number of different stages of varying abstraction levels, as exemplified in the visual system. Here, stimuli are sequentially processed from simple processes such as detection of edges and identifying primitive shapes, to understanding more complex visual shapes, before finally making sense of the visual image.

The human brain consists of approximately 86 billion neurons, also known as *nerve cells* or *brain cells*, are basic functional units of the nervous system, which carry and transmit information via electrical nerve impulses over so-called synapses (Azvedo et al., 2009; Martin, 2010a). On average, each neuron has about 7,000 synaptic connections with other neurons, with some having up to 15,000 synapses (Martin, 2010b). When a nerve impulse reaches a synapse, a neurotransmitter is released, which triggers an electrical impulse in the next neuron. This is the process which deep learning researchers have been trying to emulate for decades, and it is easy to see the similarities in deep learning systems. The following sections describe how such processes are constructed in deep learning systems.

As the name implies, *Artificial Neural Networks*, often simply referred to as *Neural Networks*, are unmistakeably related to the processes and activities of the brain. The idea of building complex intelligent systems from a large number of more simple computational units first surfaced in the 1980's, and has now been successfully applied in practice (Goodfellow, Bengio & Courville, 2016). Over the past decade, the focus of deep learning research has changed from unsupervised learning towards supervised learning, due to the increased performance of the latter. Today, supervised learning is the most common form of deep learning, and machine learning in general. (LeCun, Bengio & Hinton, 2015). An example of a supervised learning method is the use of neural networks, which is the most widely used method for deep learning today.

Neural networks

As previously mentioned, the brain consists of billions of interconnected neurons. Neural Networks (NN) are built in a similar manner, as seen in the conceptual illustration in figure 15 below. NN consist of many different connected, small, and simple processors called *nodes*, or *neurons*. Even though a NN can contain many different architectural layers of nodes, they are divided into three main layers; the *input*, *hidden layers*, and the *output* of the network (Bengio, 2009). While the input and output both consist of a single layer of nodes, the number of intermediate layers of nodes which are neither inputs nor outputs, are all considered to be part of the hidden layer. Until 2006, no system existed which could compute more than a couple of layers, thereby restricting the number of hidden layers to one or two, and consequently the limiting the complexity of the neural networks. The work on Deep Belief Networks, published by Hinton, Osindero and Teh in 2006, was considered ground-breaking, in the sense that their research enabled building deeper multi-layer neural networks (Bengio, 2009). The Deep Belief Network introduced the idea of training one layer at a time, using an unsupervised learning algorithm on each layer, an algorithm known as a Restricted Boltzmann Machine (Hinton, Osindero & Teh, 2006). Since then, the algorithms and neural networks have developed even further with greater results, but the conceptual idea remains the same.

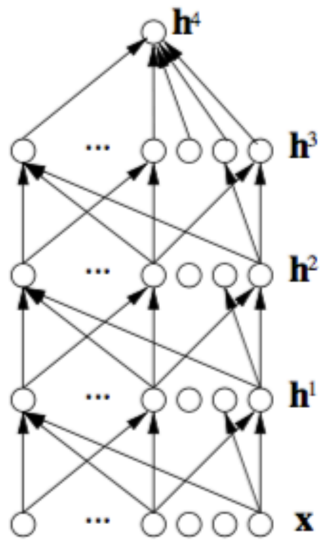


Figure 15: Illustration of a Neural Network with the input layer x, three hidden layers h1-h3, and a single output node h4.

In a NN, information can be sent from a number of different input nodes in the first hierarchical layer (indicated by *x*), to the nodes in the first hidden layer of the network (h1), whereby these nodes are activated. Based on the information received from the input nodes, and a pre-defined weighted connection, each node in h1 evaluates the input, processes the information, before finally passing its processed data on to a node in the next hierarchical layer, h2 (Schmidthuber, 2014). This process repeats through all layers, until the final output node(s) has been computed. The actual number of nodes and layers in a NN varies greatly, but some systems consist of several million nodes, and billions of connections (Pesenti, 2014). The unique and crucial element of NNs lies in the determining how much weight a given node should have in the network. These weights are determined by a process called *back propagation.*

5.4 Affective computing

An area of cognitive computing which has received a lot of attention in recent years, is that of emotional intelligence in computers, also known as *affective computing*. In her 1997 paper, Rosalind Picard, professor at MIT Media Laboratory for Perceptual Computing, first presented the term, which she defines as *"computing that relates to, arises from, or influences emotions"* (Picard, 1997, p. 1). As is the case with cognitive computing, affective computing is an interdisciplinary field, which spans across both computer science, cognitive science, social science, and psychology. Sometimes also referred to as *emotional-oriented computing*, or simply *emotional computing*, affective computing has become an emerging field of research, which aims to enable computers to recognise, feel, infer and interpret human emotions (Poria et al., 2017). The overarching end-goal of affective computing, is for humans to be able to interact and interface with machines as richly as we interact with each other (Cowie, 2005). Emotions are key components when discussing essential cognitive processes, and relates not only to human creativity and intelligence, but also plays a huge part in rational thinking and decision-making. Thus, to achieve a natural human-computer interaction, computers need the ability to recognise and express affect (Picard, 1997). Two of the most important aspects of affective computing are those of sentiment analysis and emotion recognition, as they are key components of human cognition and communication (Cambria, 2016; Poria et al., 2017).

5.5 Analysing and modelling emotions in computers

As mentioned, emotion recognition and polarity detection are the two most important aspects of sentiment analysis, and affective computing in general. The two are highly interrelated and interdependent concepts, with the former often understood and processed with emotion labels, while the latter consists of binary classification tasks, such as positive/negative or like/dislike. The interdependency often manifests itself in polarity detection being inferred from emotion recognition, and hence, emotion recognition is sometimes considered a subtask of polarity detection (Cambria, 2016). Despite its recent uprise, technologists have largely ignored taking emotion into consideration in systems design throughout the years, which has often lead to frustrating and unsatisfying user experiences. This complexity is due to the multifaceted aspect of emotions, such as cognition, conation, interaction, personality, and culture (Cowie, 2005).

Lee and Norman (2016) categorises the current computational models for modelling emotions into five overarching areas; dimensional, anatomical, rational, communicational, and appraisal. Dimensional models focus on high-level core affects (such as positive, negative, or mood) and do not categorise emotions into discrete states (such as fear or happiness). The anatomical models are built with a starting point in brain theories and neural science, and ties emotions to certain neural and biological processes. The third category, rational models, concerns the functions or roles that are facilitated by certain emotional states, whereas communication models aims to endow the behaviour of social intelligence by sharing or displaying emotional states. The most widely used theory for modelling emotions today, are appraisal models. Such models are component based, meaning that the model itself consists of several sub-models. Appraisal models are based on the appraisal theory, in which it is believed that any given emotional state is based on an individual's evaluation of the surroundings, situation, or contextual cues (Lee & Norman, 2016). Based on the emotional state, certain physical and cognitive behaviour will manifest in the person. As the situational context changes, maybe as a result of the individual's expressed behaviour, the situation is re-evaluated, leading to a recursive feedback loop which can enable different emotional states.

The following sections will discuss how emotions can be modelled in a computer. Specifically, emotions derived from facial recognition, speech processing, text analysis.

Facial expressions

Facial recognition and expressions are extremely important mediums in interpersonal communication, which allows us to express emotions or communicate conversational cues, either consciously or unconsciously (Tao & Tan, 2005). They make it possible to form an impression of another person's sentiment and emotional state (Poria et al., 2017).

A prerequisite for detecting emotions from faces in images or video, is the action of detecting the face itself. Facial recognition has been an important research area for computer scientists for many years, and it is a key element in a number of different face-related applications, including emotion recognition. Facial recognition research has seen immense improvements in recent years. In 2001, the first ground-breaking real-time facial detection system was developed by Paul Viola and Michael Jones, and since then the detection rate has greatly increased (Viola & Jones, 2001; Tao & Tan, 2005; Sun, Wu & Hoi, 2017). Throughout the years, a number of different methods have been used for facial expression recognition, such as feature extraction, Support Vector Machines, Gabor wavelets, and Hidden Markov Models (Tao & Tan, 2005). In recent years, however, deep learning and deep convolutional neural networks has achieved remarkable success in computer vision. Deep learning methods have a great advantage over traditional computer vision methods, since it greatly reduces manual engineering and hand-crafted designs (Sun, Wu & Hoi, 2017).

Since the emergence of these techniques, different neural network applications have dominated object detection benchmark evaluations, including the ImageNet Visual Recognition Challenge (ILSVRC). Sun, Wu, and Hoi (2017) extended the Faster region-based convolutional neural network by combining several different strategies to achieve state-of-the-art results on the widely used Face Detection Dataset and Benchmark (FDDB). An example of a specific application using neural networks is IBM Watson's Facial Recognition, which is described in section 5.6.

Speech processing

The spoken language is used to convey meaning, without have any physical manifestation per se, unlike e.g. image processing where an object within the image can be categorised based on its features, making speech analysis extremely difficult (Manning & Socher, 2017). The goal of natural language processing (NLP), is to allow computers to understand natural language, from which they are able to perform a given task. They range from relatively easy tasks, such as spell checking or keyword search, to harder and more complex tasks like semantic analysis and question answering (Manning & Socher, 2017). Parameters such as accents, pronunciation, articulation, pitch, and noise, makes speech recognition a highly complex process. Previously, speech recognition models were mostly built on Hidden Markov Models, but the accuracy of such models do not compare to recent results.

The capabilities of speech recognition systems have achieved breakthrough results this year, by achieving a lower word error rate (WER) than humans are capable of. WER is an indicator of the inaccuracy of recognising spoken words. Research has shown that the WER of humans is approximately 5.9%, but in May 2017, Google CEO Sundar Pichai announced that their speech recognition technology achieved a record WER of 4.9% (Protalinski, 2017). In comparison, Microsoft achieved a 6.3% WER in September, 2016, whereas IBM managed a WER of 5.5% in March, 2017 (Saon, 2017; Tung, 2016). Pichai states that Google's ground-breaking achievement is a result of the advent of neural networks when processing the speech signal (Protalinski, 2017).

Analysing emotions from speech is not only a matter of understanding the verbal content, i.e. *what* is being said, but *how* it is said can play an even bigger part in determining the actual emotional state. Features such as pitch, loudness, speaking rate, pauses, rhythm, and the correlations between them, all contribute with emotional information (Gangamohan, Kadiri & Yegnanarayana, 2016). Paralinguistic features carry a lot of affect information, and if these are not considered, important information can be lost, which can ultimately lead to a complete misunderstanding of the utterance (Pantic & Rothkrantz, 2003).

The process of emotion recognition in speech happens relatively effortlessly in interpersonal communication, despite being a highly complex phenomenon which is not well understood. As a result, it is very difficult to recreate this process in a computer (Gangamohan, Kadiri &

Yegnanarayana, 2016). In general, the process of emotion detection in speech consists of six components; speech input, speech normalisation, feature extraction, feature selection, a classifier system, and emotion detection (Sudhkar & Anil, 2016). The most important part is to extract and select the relevant features from the speech input. The features are usually divided into three categories: prosody, voice quality, and spectral features (Gangamohan, Kadiri & Yegnanarayana, 2016). An example of using prosodic features is Sudhkar and Anil (2016), who describe different basic emotions based on parameters such as pitch, intensity, speaking rate, and variance. Another highly important element of the emotion recognition process, is to create an accurate classifier, but in order to do so, a good and comprehensive database of annotated emotional speech is needed. Collecting a sufficiently extensive database is highly problematic, since many features are considered to be speaker and sound-specific, which makes classification rather difficult, as described by Gangamohan, Kadiri and Yegnanarayana (2016). They even go as far as stating that emotion recognition by a machine is an “*elusive”* goal.

Emotion-mining in text

Detecting and analysing affect from text is another widely researched field within affective computing, known as emotion-mining. Yadollahi, Shahraki and Zaiane (2017) defines four different types of emotion-mining tasks: emotion detection, emotion polarity classification, emotion classification, and emotion cause detection. Emotion-mining can be applied in a number of different areas. It can be used to derive levels of satisfaction in customer service; allow content filtering based on emotions expressed; author profiling; predicting depression, stress, or suicidal thoughts; and of course to determine emotional states in tutoring systems (Yadollahi, Shahraki & Zaiane, 2017). Emotion-mining can be performed on e.g. personal notes, emails, news articles, blogs, books, or chat messages. The emergence of social media such as Twitter and Facebook, has opened the door for new possibilities within emotion classification and social media analytics, as exemplified by Farías, Patti and Rosso (2016), who proved that affective information can be used to detect irony in tweets.

Classification of emotions can be done on either document or sentence level. When analysing on document level, an entire document of text is the unit of input, and thus the emotional tone of the whole document is derived. As the name implies, sentence level classification determines the polarity and tone of a single sentence (Yadollahi, Shahraki & Zaiane, 2017).

Cambria (2016) divides the existing approaches to emotion recognition and sentiment analysis into three main categories; knowledge based techniques, statistical methods, and hybrid approaches. In knowledge based techniques, text is classified into categories of affect, based on the presence of words that are rather unambiguous, such as “happy” or “sad”. However, these techniques have one major weakness, that is poor recognition in more complex linguistic constructs. The second category, statistical methods, works by feeding a machine learning algorithm a large training data set with affectively annotated texts, from which the computer learns e.g. the affective valence of the affect key words, as well lexical affinity and word co-occurrences (Cambria, 2016). The third category, hybrid approaches, harness and combines the power of the former two, from which it is possible to recognise emotion and detect polarity from unimodal or multimodal data. Hybrid approaches are considered more accurate, especially when classifiers trained on large data sets, analyses large text inputs. Most academic research incorporates data sets and sources of affect words such as the Affective Lexicon, SentiWordNet, and SenticNet. A number of different methods for emotion-mining exists, such as Support Vector Machines, rule based algorithms, and Hidden Markov Models. In addition, deep learning and neural networks have been used with successful results, as they are able to grasp more complex conceptual rules, and shift the approach from being syntax-based towards more semantics-aware frameworks (Cambria, 2016).

5.6 State of the art cognitive and affective systems

This section will present some of the most advanced cognitive and affective systems developed today. This includes state of the art technological solutions, different development platforms, and cognitive assistants.

DeepMind

Acquired by Google in 2014 for $400 million, DeepMind Technologies is an industry leader and a global innovator within cognitive computing. Over the past decade, the company has achieved great success in building neural nets for deep learning (Pereira et al., 2015).

DeepMind has developed a computer program with the capabilities of learning to play Atari video games from scratch, by using a reinforcement learning system based on recurrent neural networks. Using a trial-and-error approach without human intervention and guidance, the autonomous system learned and developed over time, reaching and even surpassing human-level performance (Goodfellow, Bengio, and Courville, 2016; Pereira et al., 2015). Using computer vision capabilities, the system is capable of processing and interpreting both natural language and visual data streams, and to learn and act based on the perceived data.

While Atari games are rather simple of nature, DeepMind demonstrated the true capabilities of cognitive computing, when their program *AlphaGo* defeated the World's best player of the game *Go,* Lee Sedol (Hassabis, 2016). *Go* is widely considered the most complex board game invented by mankind, and thus winning four out of a five games-series came as a huge surprise to the general public. The victory was a true eye-opener for many, as AlphaGo made moves which only had a 1 in 10,000 chance of being played by a human. Due to the complexity of the game, it would be a virtually impossible task to code all possible moves with a brute-force approach, further emphasising the ground-breaking progress currently being made within the field of cognitive computing. According to CEO and co-founder of DeepMind, Demis Hassabis, AlphaGo's ability to find solutions that humans do not learn or consider, has huge potential in other application areas and domains as well (Hassabis, 2016). Furthermore, Hassabis states that even though the match demonstrated how humans and artificial intelligence can push each other towards new ideas and solutions, there is still a long way for machines like DeepMind to be fully capable of the intellectual tasks of humans.

IBM Watson

For many years, IBM have been considered the frontrunner within the field of cognitive computing, and their cognitive technology, Watson, is a technology that has the capabilities of thinking like a human (IBM, 2017b). It helped making the capabilities and possibilities of cognitive computers become mainstream after it famously won the game show Jeopardy in 2011, after beating two of the best human contestants (Jackson, 2011). After decades of research and more than five years of development, Watson was able to understand, analyse, and answer the questions asked by the host (Kelly & Hamm, 2013). During the show, Watson did not have access to the internet, but could access four terabytes of encyclopaedia data (200 million pages of content) stored locally. Since then, the capabilities of Watson have vastly improved. It has evolved from being question-and-answer computer, to a supercomputer which can analyse and interpret all kinds of structured and unstructured data, such as text, images, audio and video. Watson can understand a user's personality, tone, and emotion, which can then be used to e.g. provide personalised recommendations, or to create emotionally intelligent virtual agents (IBM, 2017b). In 2016, IBM made it possible for developers to harness the power of Watson, by making APIs available along with the launch of the Bluemix platform, which is further elaborated in section 6.3, Bluemix. It also powers one of the most powerful data analytics platforms on the market today, *Watson Analytics* (IBM, 2017c).

Case: Watson for Oncology

Today, Watson is used in a vast amount of different applications across all sectors, with the Watson for Oncology being a prime example of its capabilities and potential. Based on expert training from physicians at Memorial Sloan Kettering Cancer Center (MSK), Watson provides clinicians with evidence-based treatment options for cancer patients (IBM, 2017d). When applied to a new domain, it is essential that Watson learns the language and terms used in that specific domain. This is done by feeding Watson huge amounts of structured and unstructured data, in this case this includes physicians' notes, patient information, literature of the best medical science within the field of oncology, as well as medical reports of patients with a similar diagnosis (IBM, 2014a). This data is then indexed, and Watson is trained with question/answer pairs which serves as ground truth, and even though it does not give explicit answers to all possible questions, it helps Watson understand the linguistic patters in oncology. Based on this training, Watson is able to understand inquiries from the doctors, from which it builds and evaluates hypothesis. Using statistical data modelling and weighted evidence scores, Watson estimates its confidence of a hypothesis based on this evaluation. This ability to instantaneously run analytics on a vast body of data, and evaluate each unique patient against clinical evidence, allows the oncologists to make better and more informed decisions (IBM, 2017d; IBM, 2014a).

Development platforms

In recent years, tech companies have sought to change or expand their business model, by allowing developers access to their cognitive technologies (Pereira et al., 2015). Market leaders and tech giants IBM, Microsoft, Qualcomm and Hewlett Packard are examples of such companies, each with their own development platforms (Google, 2017; Hewlett Packard Enterprises, 2017; IBM, 2017b; Microsoft, 2016; Zaki, 2015). These platforms allow developers and businesses of any size to use some of the most advanced cognitive technologies that exist today. Access to these services can become powerful catalysts across all industries (Pereira et al., 2015). Visual recognition, face detection, speech analysis, and language translation are some of the services available. The following sections presents two of the leading cognitive development platforms in the market today, IBM's Bluemix and Microsoft's Cognitive Services.

IBM Bluemix

After 18 months of development, IBM released its Bluemix platform in February, 2014 (IBM, 2014b). Bluemix allows developers to build web applications, using a number of different services in the areas of data and analytics, storage, network, and commerce. Some of the most noteworthy services on the platform are those provided and powered by Watson, IBMs cognitive super computer. These services include speech-to-text and text-to-speech conversion, conversational services, customisable visual recognition, as well as language classification and translation (IBM, 2017e). Furthermore, services such as *Tradeoff Analytics,* helps people make more informed and better choices, by taking into account multiple and sometimes conflicting goals (IBM, 2017f). This is a great example of an emulation of how humans think, behave, and make purchasing decisions, where multiple criteria are usually taken into consideration. In real estate, for example, compromises must often be made in one or more factors such as price, mortgage, location, and size. Tradeoff analytics can help people making such a decision. In relation to affective computing, two of the most interesting services provided by Watson are *Personality Insights* and *Tone Analyser,* both of which will be described in more detail in section 6.3, along with a number of other Watson services.

Microsoft Cognitive Services

Previously know as *Project Oxford,* Microsoft released their *Cognitive Services* platform in 2016. Microsoft categorises their cognitive services into four different categories: vision, speech, language, knowledge, and search. The vision services extract rich information from images and video, and provides APIs for e.g. facial detection, intelligent video processing, content moderation, and emotion analysis. All the speech processing and language services are powered by Microsoft's *Language Understanding Intelligent Service* (LUIS), which allows for speech, speaker, and intent detection, as well as conversions back and forth between text and audio. Other services, such as linguistic analysis are also available. Using active machine learning, the precision and accuracy of LUIS improves over time, thereby continuously enhancing the capabilities of the service (Microsoft, 2017a).

In addition to vision and speech capabilities, Microsoft also provides knowledge services, which can be used to e.g. understand and distinguish different meanings of a word depending on the context in which it is used. Even within the same paragraph, the so called *Entity Linking Intelligence Service*, can identify the different entities. Furthermore, a number of intelligent search APIs are available, applicable in both image, news, video, and web search. Microsoft are continuously making new services available, to further increase the already impressive API catalogue (Foley, 2017).

Cognitive assistants

Today, *Cognitive Assistants* (CAs), sometimes also referred to as *virtual assistants* or *intelligent assistants,* are an integral part of how people interact with their phones, computers, and other electronic devices. According to Hamid Motahari, research group lead on IBM Cognitive Services, CAs offer *"computational intelligence capabilities typically based on natural language processing, machine learning and reasoning, and provide cognition powers that augment and scale human intelligence"* (Motahari, 2013). The emergence of CAs have made it possible to perform tasks such as ordering a taxi, scheduling meetings, performing searches, and assisting in navigation, simply by issuing a voice command or asking a question. CAs are designed to understand natural language inputs, interpret and process the request, and respond with an appropriate answer in natural human language (Waters, 2016). Today, CAs are only partially based on its user, with personal data such as calendars and addresses, but in the future, the CAs might even know more about a person than the person himself, and the technology will develop from being separate from its owner, towards the two being embodied and unified as one (Bauer, 2014). Leading technology companies have developed their own IAs, with Google Assistant, Apple's Siri, Microsoft's Cortana, and Amazon's Alexa being the most prominent and widely used CAs. As an example of the popularity of these services, Siri alone received more than two billion requests per week in October 2016, according to Apple CEO, Tim Cook (Clover, 2016). CA is an area in which the tech giants are currently fighting a battle of acquiring new technology, and making their product more sophisticated and superior to that of their competitors, as there is money to be made in automating everyday processes (Waters, 2016; Corbyn, 2016).

Oren Etzioni, head of Allen Institute for Artificial Intelligence, states that even though CAs have had a major breakthrough, today's devices do not compare to the kind of capabilities that is wanted and envisioned by the technology sector (Corbyn, 2016). He sees the future market leader, as the company that can make the CA act more like a hotel concierge, with sophisticated dialogue and the ability to handle far more complex tasks, better and faster than possible today in the above mentioned technologies. A technology that aims to overcome this, is Viv. Acquired by Samsung in 2016, Viv is an open artificial intelligence platform whose goal is to provide consumers with a single intelligent conversational interface to interacting with devices and services everywhere, by allowing developers to integrate their services and capabilities to the platform (Viv, 2017; Samsung, 2016). Instead of hardcoding all possible outcomes to a request, Viv detects the user's intent, and generates the code dynamically using patented Dynamic Program Generation (Kittlaus, 2016).

Other application areas

In the marketing domain, cognitive computing has also seen a prominent rise and is widely applied throughout the industry. A great example of a cognitive technology, is Equals 3's *Lucy*. Powered by IBM's Watson, Lucy is a software which can understand natural language questions and inquiries, and based on big data sets of structured and unstructured data, Lucy can analyse and present corresponding results in a visually appealing interface (Equals 3, 2016). It can help marketers by building audience insights, and perform market and segmentation analyses, and automatically visualise and describe the most relevant findings from the data.

Another application area in which affective computing has long been researched is within the field of robotics. Throughout the years, many different emotionally intelligent robots have been developed, each with their own individual purpose and characteristic, but they all have the common trait of understanding and/or expressing emotions. The most widely known emotionally intelligent robot today, is the humanoid robot, *Pepper*, developed by Softbank and Aldebaran Robotics. Using directional microphones and high-definition cameras, Pepper can identify emotions by analysing speech (e.g. tones and lexical use), facial expressions, and posture, and communicate in a corresponding tone (SoftBank Robotics, 2017). Today, Pepper is used in a number of different application areas, ranging from in-store customer service, to teaching assistance at universities.

5.7 Chapter summary

This chapter investigated cognitive computing, i.e. the simulation of human thought processes in a computerised model. Cognitive systems are capable of interpreting complex information, reason and act upon that information, and even develop and learn over time. The recent progress in cognitive computing, is a result of developments within a number of fields, such as pattern recognition, machine learning, big data, and cloud computing. To clarify the functionality of cognitive computing systems, deep learning and neural networks were also presented.

To properly research technologies related to affective tutoring systems, the field of affective computing was also investigated. Affective computing is an area of cognitive computing concerning computing that relates to, arises from, or influences emotion. Different aspects of affective computing were discussed, including emotion recognition from facial expressions and speech signals, and emotion-mining in text. Finally, different state of the art cognitive and affective systems and technologies were discussed, such as DeepMind and IBM Watson, different cognitive development platforms, and a number of cognitive assistants.

6. Analysis

After investigating relevant research on the topics of interpersonal communication, human emotions, and cognitive and affective computing, it is possible to evaluate the capabilities of an ATS built to emulate human tutoring. This chapter will first and foremost present a conceptual ATS model, before discussing the extent to which it can actually replicate the interaction taking place within interpersonal communication between a student and a teacher. Furthermore, a secondary model will be presented, illustrating how different cognitive technologies can be utilised in a small-scale ATS implementation. The purpose and functionality of these models are then discussed with the four university students, whom where also interviewed in section 3.2, to gain their view upon such a system's viability, value creation, and potential application in their daily learning environment.

6.1 Affective Tutoring System Model

This section will present the conceptual ATS model, in terms of both the requirements for the system, its overall structure, each separate component, and the individual functionalities of the components.

Purpose and functionality

The main purpose of the ATS is to communicate with the student in a manner which emulates interpersonal communication. In order to appropriately do so, the ATS must consider a number of factors in relation to both the student, the verbal and nonverbal messages received, and take all these factors into consideration when constructing the response. More specifically the tutoring assistant must:

1. **Understand the student**: Understand the student as a person, as a learner, and as a communicator. As described in the framework, this includes a wide variety of factors such as personality, social and cultural contexts, psychological and physiological factors, and the internal states of the student.
2. **Understand the message**: Interpret and understand the message that is communicated by the student. The main concern here is to decode the message encoded by the student, mostly through verbal and nonverbal cues.
3. **Respond appropriately**: Act appropriately during the interaction process, construct a response, and communicate according to the message and emotional state of the student.

The following section will present the conceptual idea, before going more in-depth with these three aspects, and present different cognitive services which can be used to build an affective tutoring system.

Conceptual model

Based on the research from chapters 3, 4, and 5, as well as the system requirements presented above, a conceptual idea model is formed, describing the architecture of an ATS. An illustration of the model can be seen in figure 16. In this section, each of the different model components and their functionality will be described. The five main components are: the student as a communicator, the interface in which the messages are sent, background information on the student, the emotion processor, and a pedagogical unit.

Student

The first component of the model is the student himself. In this component, the student is considered a communicator, i.e. simultaneously sender and receiver. The most important elements are how he interacts with the system, and how the feedback from the system is received. Aspects such as individual characteristics and personality is not considered at this stage in the model.

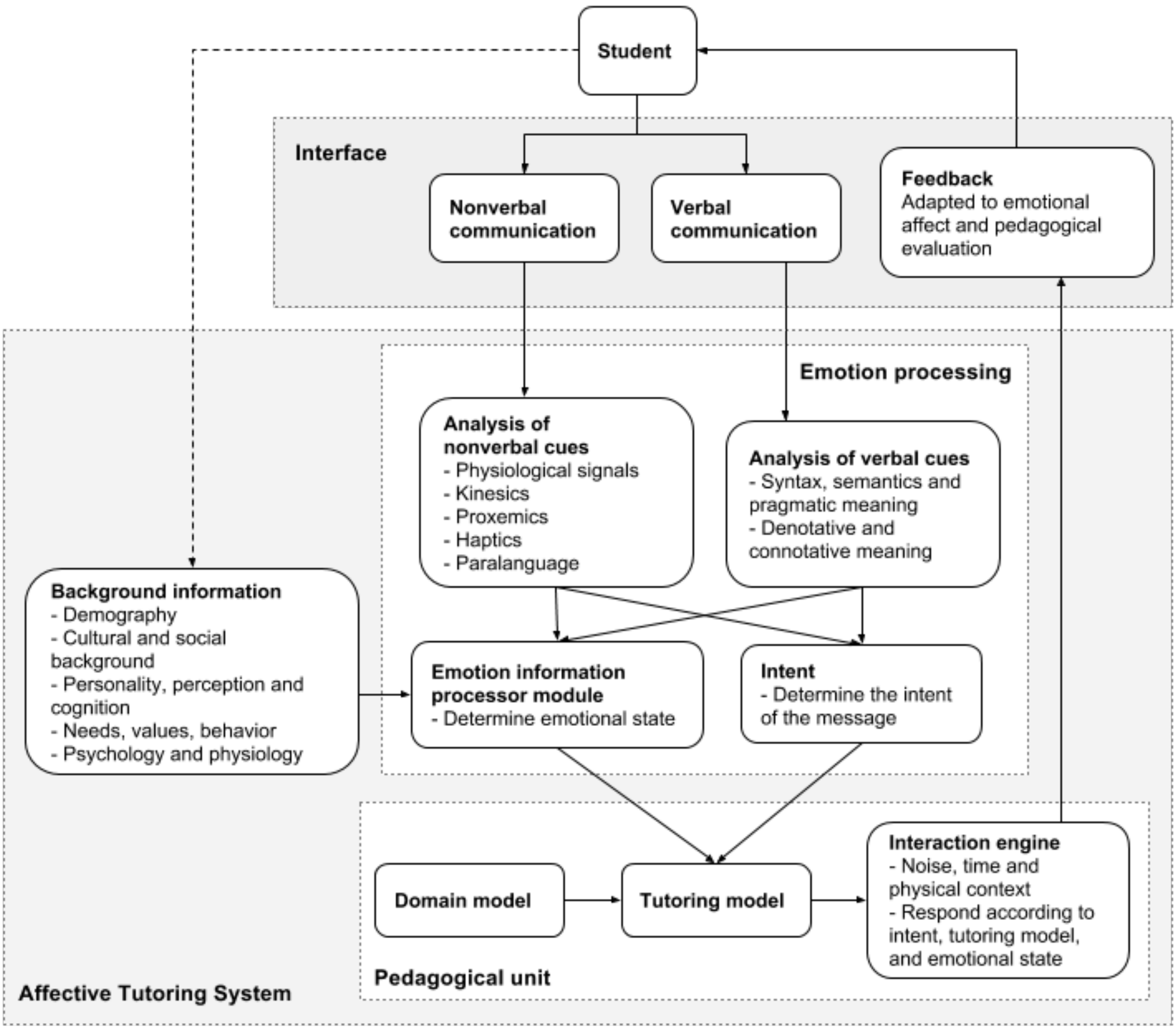


Figure 16: The conceptual ATS model.

Interface

The second component, the interface, is where the actual messages are transmitted. This relates to both verbal and nonverbal communication, from both the student and the ATS. While much information can be drawn from **verbal communication**, **nonverbal communication** signals are of extremely high importance when discussing communication with affect, and must also be incorporated in the model. It is also in the interface, that the adapted **feedback** constructed by the ATS is sent to the student.

Background information

The first part of what is considered the core ATS, is the background information component, which contains valuable information about the student. Some of the most relevant information includes demographical data, and social and cultural background. Such rather simple information can be of great value in the big picture of communicating in a personal manner. This type of data can be used in later stages of the communication process, such as when interpreting a message and constructing the response. Another important aspect to consider is the student's personality. Individual needs, values, and behavioural traits, are key elements in understanding the individual student. As mentioned in section 5.4, this information is also important when determining the pedagogical strategy in the tutoring model. Finally, both physiological and psychological characteristics should also be included.

Emotion processing

The second part of the ATS, and the fourth component of the overall model, is the emotion processing component. As the name implies, this is where affect information captured in the interface is analysed. Through **analysis of verbal and nonverbal cues**, the message and emotion signals are evaluated. The results from this evaluation serves two purposes. First, the emotion tones are sent to the emotion information processor, and secondly, the student's **intent** with the message is evaluated.

What separates ATS from ITS, is the inclusion of affect data, and subsequently an **emotion information processor module**. This is an essential part of the ATS, as this is where all the affect information is evaluated. Besides the emotional tones derived from verbal and nonverbal cues, the student information must also be considered, as individual student characteristics carries highly valuable information. The overall emotional state of the student, is determined using a custom algorithm that weighs all the inputs, and computes a single output, which can then be used in the tutoring model.

Pedagogical unit

The fifth and final component of the model is the pedagogical unit. In this component, the learning strategy is evaluated, and the feedback to the student is constructed. The first element in this component is the **domain model**. As described in section 3.1, the domain model contains the relevant information and knowledge within a certain learning domain. By comparing student questions or responses to this information, it is possible to evaluate the relevance and accuracy hereof, before using it as an input to the tutoring model. The **tutoring model** process this information along with the student information and the emotional state computed in the emotion information processor module, and determines an appropriate pedagogical strategy to use.

The third element within the pedagogical unit is the **interaction engine**. As described in section 4.4, the interaction is considered the core component of interpersonal communication, which is also reflected in its complexity within the system. This is the component in which all information is combined, and the final response is constructed and sent to the student via the interface.

6.2 Replacing human tutoring

This section evaluates the individual elements within the ATS model, and the extent to which the proposed solution can emulate personal tutoring, as presented in the interpersonal communication model. This part of the analysis will be made with a starting point in the three components of the ATS, as presented in the previous section, and thus concerns the following three areas; information about the student, the processing of emotional display, and constructing the response in the pedagogical unit. In other words, understanding the student, interpreting the message, and responding appropriately.

Understanding the student as a communicator

Whether communicating with a teacher or interacting with an ATS, the individual characteristics of the student himself do not change. These characteristics, however, remain just as important when it comes to the ATS interpreting the messages from the student, as it is in interpersonal communication. The student remains the same unique individual, in regards to all of the factors mentioned in the communication model: psychological and physiological factors; perception, cognition, and personality; interpersonal skills and competencies; as well as the cultural context and demography. Since each of these aspects help define the student as a communicator, they must all be considered and implemented as thoroughly as possible in the ATS. Due to the nature of the interpersonal communication, the individual characteristics must be taking into consideration when both decoding the messages sent *from* the student, and when constructing a message *to* the student. In the ATS model, the background information module stores most of the student information mentioned in this section. The following subsections will discuss various psychological, physiological, and social factors to consider within the ATS.

Perception, cognition, and personality

Other aspects related to the psychology of the student, are his perception and cognitive processes, as well as the characteristics of his personality. These aspects are all complex mental states and processes, and are thus very hard to assess in a straightforward manner by the ATS, although it is possible by analysing and building a mental profile of each individual student. This could address traits such as needs and values, information and memory processing, social cognition, as well as the attitudes, motivation, goals and intents of the student. Such information would be important additions to the student information module, and the subsequently related processes of emotion processing, evaluating intent, and deciding on the tutoring strategy and response. Furthermore, personality traits such as affective orientation, assertiveness, self-esteem, and communication style could also provide valuable information within the ATS. A number of tools and tests exists for evaluation of both perception, cognition and personality traits, and by implementing such analysis results, the learning effectiveness could improve. A practical example of a personality analysis tool is described in section 6.3.

Social and Interpersonal skills

In interpersonal communication, social and interpersonal skills and competencies play a huge role in how we interact and communicate with others. Factors such as social perspective and role taking, trust, and social competences, defines actions and roles depending on the social contexts of the interaction. This especially applies to self-disclosure within interpersonal relationships, deceptive behaviours and detection of deception, the use of compliance gaining strategies, and detection and resistance to other people's compliance gaining strategies.

Despite their importance in interpersonal communication, it can be argued that several of the mentioned factors do not directly apply within human-computer interaction. To maximise learning effectiveness, however, the ATS should consider e.g. deceptive behaviour and behaviours associated with self-disclosure, in order to fully comprehend the actual skill level and learning speed of the student, as well as analysing the student's affect display correctly. Detecting these signals, is one aspect where human tutoring is often superior to the ATS, and thus it must be implemented within the system to optimise the effectiveness hereof. The research conducted throughout this project did not reveal any existing ATS with such capabilities, highlighting the potential improvements that lies within this area.

Cultural context and demography

In transactional communication models, and thus interpersonal communication, culture is considered an overarching component, affecting all elements of the communication process (Stamp, 1999). In the ATS model, culture and cultural context are mainly concerned with the student and the message composed by the student. Cultural aspects are important considerations in both understanding the student as an individual, and in responding with appropriate feedback.

Communication varies, and some signals are perceived differently within different cultures, thus, the student's cultural background must be properly understood in order to decode the messages correctly. In other words, two students with different cultural background, might express different messages with the same intent, or interpret the same message in different ways. Amongst others, the cultural context concerns information acquisition, relationship types and terms, and communication problems and constraints. Whereas e.g. relationship types and intimacy levels are difficult to apply to an ATS, demographical information such as age, gender, educational level, social status, and even ethnicity, is easily implemented and could provide valuable information about the student. Due to ethical concerns, however, this could be considered personal data, and thus not applicable within the system.

While many of these aspects and traits mentioned are very difficult to evaluate by the ATS, they are also extremely difficult for people to analyse and take into consideration in a communication context. It has been proven that the superiority of individual human tutoring is a due to the ability of affect recognition from more complex signals and cues, as well as their personal knowledge about the student and his needs (Erez & Isen, 2002; Mao & Li, 2010). Some characteristics, however, are more straightforward and easier evaluated than others. As an example, demographic information can be accurately described within the ATS as simple key-value pairs (e.g. "age = 21"), whereas social identity, personality, and cultural context are much more complicated elements to handle within an ATS, due to their multi-faceted nature.

Interpreting the student's message

The second main functionality of the ATS is to interpret the communication signals from the student. In the communication model, four different elements exist within the message component: verbal cues; nonverbal cues; apprehension and encoding; and the message type. To some extent, these are all addressed in the proposed ATS model, as will be discussed in the following paragraphs.

Analysing verbal and nonverbal cues

By analysing both verbal and nonverbal cues from the student, much more complex information can be derived, compared to only using one or the other, and as a result, the accuracy of affect recognition improves drastically (Fujishin, 2008; Lane, 2008). Verbal cues can be analysed with natural language processing, from which the denotative and connotative meaning can be derived, based on the syntax, semantics, and pragmatic rules. Today's speech recognition technologies are extremely adept in regards to transcription and understanding what is spoken, and some systems even learns from the speaker and improves over time. When it comes to the actual meaning of the verbal content, additional factors must be taken into consideration, such as the accompanying nonverbal cues and contextual information. By analysing different modalities, a wide range of nonverbal cues can be evaluated within the ATS. Visuals are considered to be the one of the most dominant input in affect recognition, and especially facial expressions carry a great amount of affect information (Lane, 2008). Along with gesture and posture recognition, facial expressions are analysed using image processing and advanced machine learning algorithms. These cues are important parts of determining both the emotional tone, and the intents of the message, also referred to as the message type and subsequent message behaviour. While technological solutions capable of detecting emotional tones exists (see section 6.3), it is extremely difficult for computers to detect motives within a message with the intention of e.g. ridicule or sarcasm, especially in systems using few modalities. This is one area in which the technology is currently not as capable as its human counterpart. Similar sentences can be constructed, but depending on the tone with which the message is communicated, the intention can greatly differ. While paralinguistic cues can be derived from speech, the results from analysing prosodic features are still some way off of human capabilities (Gangamohan, Kadiri & Yegnanarayana, 2016).

Evaluating apprehension and encoding

Another complex part of understanding the student's message, is to gain insights into his apprehension and message encoding. The encoding is a result of behaviours associated with apprehension, i.e. how the student perceives the receiver's expectations and apprehension, as well as how different types of message encoding are perceived by the receiver. To proper understand this thought process, a deeper understanding of the student's personality and interpersonal competences is required, as well as information about the student's cultural background, and other individual traits is relevant. Thus, the accuracy in decoding the message depends partially on the level of detail within the student information component, as presented in the previous section.

Determining the overall emotional state

Besides their roles in understanding the intents of the student, analysing the verbal and nonverbal cues are key elements in determining the emotional state. In interpersonal communication, this is largely a subconscious process, but in affective computing systems, this must be implemented using advanced algorithms. The results from each individual modality is weighted and evaluated according to all other affect information. This also includes using background and personality data, which serves as a ground truth, from which the emotional deviations can be calculated, and the dominating emotional state is determined. In the ATS model, this evaluation happens in the emotion information processor module.

Responding appropriately

After analysing the student's message and emotional state, the ATS must respond according to the affect display and the student's intent, and doing so requires a number of different processes to be carried out, as described in the following paragraphs.

Functionality of the tutoring model

In the ATS model, the response is decided in the tutoring model, which draws on information from the emotion information processing module, the identified intent, and the domain model. It can be argued that the domain model acts as a contextual identifier, which is then used to determine the relevant contextual response. Together with the domain model, the intent and content of the message determines what elements the response should contain, i.e. *what* the ATS responds. The equally important aspect of *how* it responds, is determined from the evaluation of the student's emotional state, based on the output from the emotion information processing module. Thus, the corresponding decision making process in interpersonal communication, where the teacher decides what and how to respond in order to motivate and challenge the student, is implemented within this unit in the ATS.

Communicating the response

The next step is then to actually communicate the constructed response. Whereas humans are limited mainly by their physiological capabilities, an ATS can only communicate through the channels enabled by the system implementation. While the different output modalities are not explicitly described in the conceptual model, their use is determined by the tutoring model, but enabled by the interaction engine. A typical ATS responds to the student using text, speech, or animation, but it varies depending on the exact purpose and functionality of a given system. It can be difficult to address specifically which cues to use, but this system is built around a natural speech interface, for as seamless and natural an interaction as possible.

By replacing the teacher with a computer system, the characteristics of the second communicator changes radically compared to interpersonal communication. Just to name a few of the changes: the personality is no longer shaped by needs and values, there are no affecting physiological or social factors, and demography and cultural background are non-existent factors. By utilising deep learning and enabling the ATS to learn from its interactions, it can be argued that factors such as apprehension, compliance gaining, and interpersonal skills develop over time, and as a result, the dyadic relationship between the student and the ATS is established and evolves. The development of the relationship is extremely important in regards to e.g. affect recognition and communication style. Furthermore, the compatibility of communication styles, and personalities has a huge effect on the student's motivation and engagement, and thus the learning outcome.

Another important aspect to consider when responding, is the interaction itself. As described in the communication model, the nature of the interaction changes over time. From the initial interaction, a dyadic relationship is formed between the communicators, which changes as conversations unfold, and conflicts are managed and resolved. The relationship evolves over time, into a closer relationship with certain patterns of communication, and more expressive conversations. This type of change might also be seen when interacting with the ATS, if the system is capable of taking into account previous interactions with the student. This can actually be considered an essential part in relation to the learning process, as the student's knowledge and skill is continuously monitored. By enabling the system to communicate with emotion and understand the emotions expressed, the student might develop a more personal relationship with the ATS, similar to that of interpersonal interaction.

The changing role of feedback

The final component from the communication model is the feedback. When the ATS have responded to a question or request from the student, it can evaluate the feedback from the student, by continuously analysing the communication signals. This enables the system to evaluate how messages are received, and adapt if the response does not correspond to the intended reaction. By utilising the feedback from the interaction, the ATS can more accurately adapt to each individual student profile, and as a result craft more appropriate and personalised responses. This use of feedback resembles that of its counterpart within interpersonal communication, however, within the ATS, this process is only one-way. The student does not get the same immediate feedback signals from the ATS, as he would from a human tutor, where additional nonverbal cues are available, such as kinesics, haptics and proxemics.

The discussion in this section has highlighted some of the differences and similarities between the elements in the ATS model and the interpersonal communication process. For a more practical evaluation of how the ATS can apply cognitive technologies, a prototype implementation will be described in the following section.

6.3 Implementation proposal

To evaluate a more concrete example of the proposed ATS, and small-scale prototype will be proposed. This section discusses how such a prototype can be constructed, and how existing cognitive services can be used in the implementation.

System overview

In section 6.1, a complete ATS architectural model was presented, which is used as the foundation for this proposed prototype implementation. By simplifying and adjusting some of the existing components, as well as adding new functional components, a new model is created, which can be seen in figure 17. Thus, this model does not represent a complete system architecture. Consequently, some aspects have been simplified or left out compared to the ATS model.

The two elements *analysing verbal cues* and *analysing nonverbal cues* in the conceptual model contained a variety of different cues which could be used for emotion processing. In this case, only two cues are used; facial expressions and speech, which replaces the cue components in the model. **Facial expressions** can be analysed through image processing, as enabled by Microsoft's Emotion API. On the other hand, the speech signal must be converted before being further processed. This is illustrated with two separate components, **speech conversion** and **analysing verbal content**, as two different cognitive services are used for this purpose; Text to speech, and Tone Analyser, respectively. Furthermore, the *Student Information* component is split into two different elements, **Background information** and **Student personality**. The Watson service *Personality Insights* analyses personality traits of the student, thus removing the related elements in the initial background information component. The content of the domain model and the tutoring strategies will now be implemented in the **interaction engine**, powered by Watson's *Conversation* service. This component takes both the emotional state and the converted speech as an input, from which it can detect the intent and related entities, thus eliminating the intent element in the emotion processing component. Finally, a **speech synthesis** element is included, which syntheses the output from the interaction engine, thus completing the speech interface between the student and the ATS. Elaborate descriptions of these services are made in the following section.

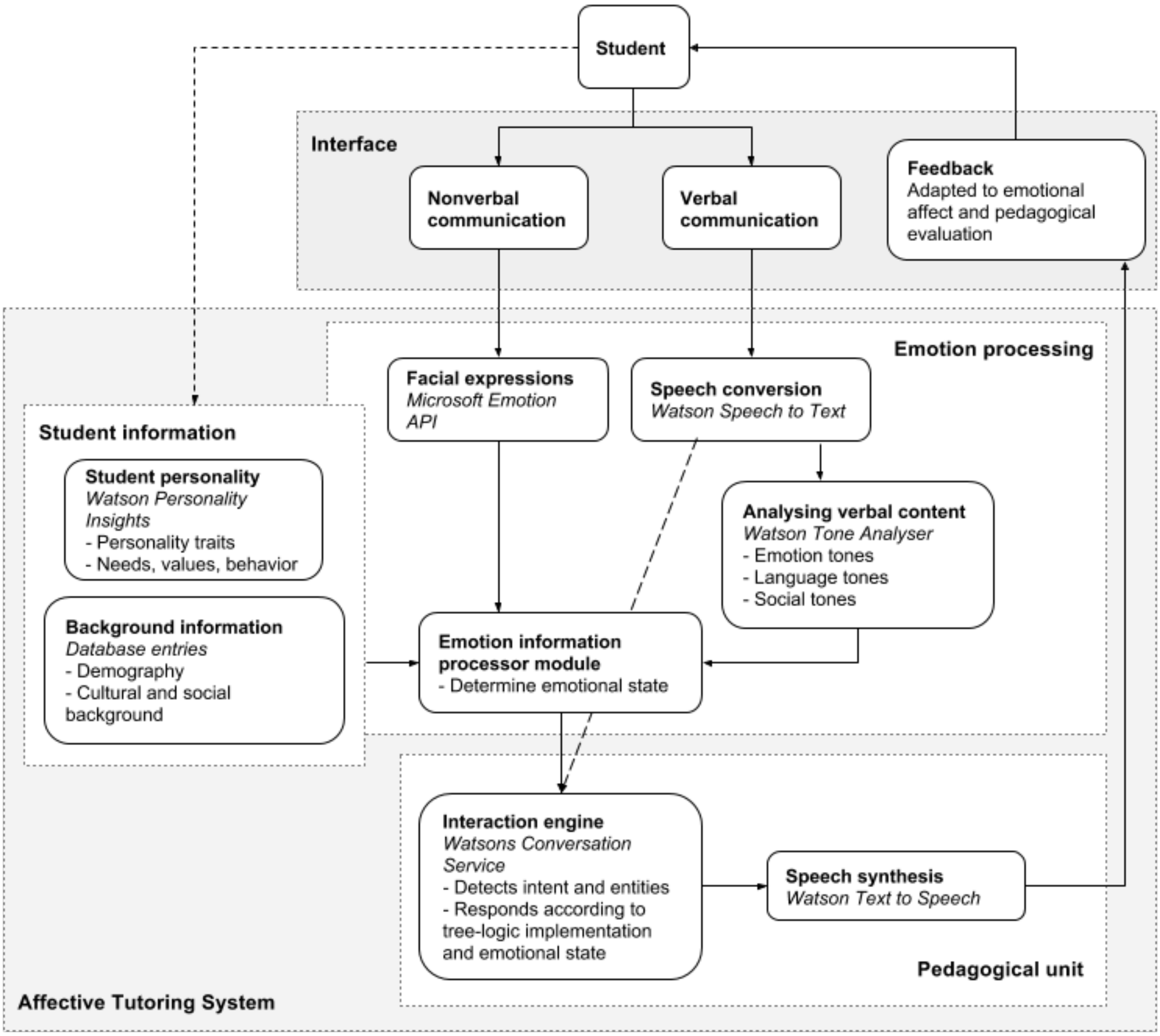


Figure 17: Illustration of the interrelations between the components and cognitive services used in the prototype model.

Use of cognitive services

This section will describe how the different cognitive services are used in the of the prototype model, covering the *Emotion API, Speech to Text, Tone Analyser, Personality Insights, Conversation,* and *Text to Speech* services.

Microsoft Emotion API

As repeatedly mentioned throughout chapter 4, nonverbal cues are extremely important components of interpersonal communication, and especially facial expressions carry great affect information. Through their Cognitive Services platform, Microsoft provides a facial recognition service called Emotion API, which can detect emotions of people in images and video. It takes an image or video containing one or more faces as an input, and returns a confidence for the emotions expressed by each identified faces in the image (Microsoft, 2017b). The dominant emotion detected in the face, is considered to be the one with the highest confidence score within the defined categories; anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

In this project, the service serves different functions. First and foremost, it enhances the tutoring assistant's ability to correctly understand what is being communicated, by improving the emotional aspect of the communication and interaction. By measuring the emotion over time, however, new rich information can be extracted from the interaction. This relates to both a change of emotional state over time, but can also be used to take feedback into consideration.

Watson Speech to Text

One of the ideas behind using a speech to text conversion in the system, is to allow the communicator to interact with the assistant as naturally as possible, i.e. via a speech interface. For the computer to be able to process and evaluate the speech input, it must first be converted to text. This can be done using the Speech to Text service powered by Watson, which is one of many services enabling speech transcription capabilities. The input signal is processed based on grammar and language structures, and is continuously re-evaluated, which means that the transcription is updated and corrected as more speech is heard (IBM, 2017g). The service allows for processing of an audio-stream, thereby enabling real-time capabilities for the system.

Watson Tone Analyser

As noted in section 4.5, interpreting a spoken message not only concerns *what* is communicated, but also *how* it is communicated. To some extent, this can be evaluated by Watson's Tone Analyser service. The emotional tone analysis is not performed on paralinguistic features, but instead conducted on the verbal cues from the written text. Using cognitive linguistic analysis, Watson can identify a variety of tones at both sentence and document level, from which the service evaluates emotions, social tendencies, and writing style of the author (IBM, 2017h).

The emotional tones are the different types of emotions and feelings that are expressed in the input. The emotions of joy, fear, sadness, disgust, and anger are evaluated according to the likelihood of them being perceived in the content. Similar to the personality analysis, the social tones are evaluated using the Big Five personality traits, which measures the person's social tendencies. This is an extremely interesting functionality, as it enables a comparison between the defined baseline of the communicator, and the current tendency expressed in the utterance. Finally, the language tones describe the perceived writing style, in regards to analytical attitude and reasoning, confidence and certainty, and tentativeness in the tone (IBM, 2017i).

In addition to the three different types of tones, the service also includes a Customer Engagement model, which allows for conversation monitoring in regards to sadness, frustration, satisfaction, excitement, politeness, impoliteness, and sympathy (IBM, 2017i). The intended use of this model is within the context of customer service and support, but can easily be applied in a more general conversational context. The insights from the Tone Analyser can be used to refine and improve communications, and craft appropriate, considerate and affective response messages to the communicator. As such, it is very applicable within an ATS.

Watson Personality Insights

To derive a detailed overview of the student's individual personality traits, Watson's Personality Insights service can be used. Personality Insights is one of the most interesting cognitive services provided by Watson. Based on a number of psychometric analyses, Watson analyses a body of written text, and returns a comprehensive personality profile for the author of the input, in this case the student (IBM, 2017j). The analysis is performed within three different areas; Big Five, Needs, and Values, and based on these results, consumption preferences of the person are inferred.

Big Five describes how a person generally interacts and engages with the world, is one of the most renowned models for evaluating personality characteristics. As the name implies, the model consists of five primary dimensions, each with a number of additional facets (IBM, 2017k):

1) Agreeableness: Describes tendencies to be compassionate and cooperative toward others, by taking into consideration altruism, corporation, modesty, morality, sympathy, and trust.
2) Conscientiousness: The tendency to act in an organised or thoughtful way. Facets include achievement striving, orderliness, and self-discipline.
3) Extraversion: Describes a person's tendencies to seek stimulation in the company of others, relating to e.g. how energetic, outgoing, and sociable the person is.
4) Emotional range: How emotions are sensitive to the individual's environment. Considers facets such as anger, anxiety, and self-consciousness.
5) Openness: The extent to which a person is open to experience within different activities. Facets include adventurousness, artistic interests, imagination, and intellect.

The second dimension of the personality analysis is concerned with the person's needs. A total of 12 different needs are analysed: Excitement, harmony, curiosity, ideal, closeness, self-expression, liberty, love, practicality, stability, challenge, and structure. The third element describes values, i.e. the motivating factors which influence the author's decision-making. The values considered by the service are: self-transcendence, conservation, hedonism, self-enhancement, and excitement (IBM, 2017k).

The insights inferred from the analysis can help understand how a person interacts with the world, how he resonates with his surroundings, and the motivating factors influencing his decision-making (IBM, 2017j). It can help determine behavioural traits and intents for the analysed person.

Watson Conversation

Depending on the characteristics of the student and the analysed message, the response must be appropriate and relevant, and communicated through a natural speech interface. The most comprehensive and user-friendly service available today, capable of handling the complexity of the interaction between a student and an ATS, is the IBM Watson Conversation service. It acts as the brain of an interaction, and allows for creation of powerful cognitive conversational applications. The service understands natural-language text inputs and, using a variety of machine learning techniques, returns an appropriate response (IBM, 2017l).

The first step of using the conversation service, is to create a workspace containing the dialog flow and training data. The dialog is defined based on so-called *intents* and *entities*. Intents represent the purpose of the user's input, such as a question about the learning material. By providing additional example key-words for a given intent, the accuracy of intent detection increases (IBM, 2017l). Based on the detected intent, a logic tree is used to navigate the dialog, which is constructed by the developer, who defines both the underlying logic and the corresponding answers to each intent. Entities, on the other hand, represents a term or an object that is relevant and related to the intents, and provides a specific context for an intent. An example of such an entity could be a specific course, book, or topic.

What really sets this conversation service apart from its competitors, is the fact that it allows for logic based on additional parameters, such as emotional states or personal data. This means that e.g. the emotional state detected by the facial expression analysis or the tone analyser, can be incorporated into the logic, from which it is possible to construct different responses, according to the emotion displayed by the student. The response could also be modified based on the gender of the student, or even on the basis of physiological signals.

Watson Text to Speech

To complete the interaction and make it as natural and human-like as possible, the constructed response must be communicated in speech, just as the inputs from the student are in spoken words. As previously mentioned, many different conversion tools and services exists, such as the Text to Speech service powered by Watson. Using IBM's speech-synthesis capabilities, the service converts written text to natural-sounding speech with minimal delay (IBM, 2017m). The output is synthesised based on a number of different models, such as acoustic models and prosodic features. Using deep neural networks, it is possible to generate values for prosodic aspects of the speech, such as duration and intonation. This is done based on linguistic attributes such as part of speech, lexical stress, positional features, and word-level prominence (IBM, 2017n). Before generating the final output, the text is transformed to a language driven by certain linguistic rules, such as abbreviation expansion, before generating the final pronunciation.

6.4 Follow-up interviews

This section will present the thoughts behind and the results of four conducted interviews, conducted on the same four students who were initially interviewed as part of the background research. They will be presented to the ATS model, from which the concept will be discussed.

Purpose

The purpose of these interviews is to evaluate the viability of the proposed concept, and to discuss the value creation of this solution. More specifically, the interviews aim to elaborate on the following key areas:

1) The students' experiences with and attitudes towards cognitive systems and personalised services.
2) The viability, value creation, and potential use of the proposed ATS.
3) The attitudes towards incorporating affect data and using of background information

As in the preliminary interview, these interviews will be semi-structured, based on a number of pre-defined open-ended questions, and the interview process itself will be an open conversation. Based on the responses from the interviews, a general discussion of the solution will be made.

Interview guide

Before doing the actual interview, the students will be presented to the conceptual idea and model, to gain a proper understanding of the functionality of the ATS. To fully cover the focus areas listed above, the following questions are constructed and used as a starting point for the interview:

1) What are your experiences with using personalised systems or services powered by AI or cognitive computing?
2) In which areas do you see the largest applicability of an ATS within a learning context?
3) How do you see the ATS being used to support learning and answering questions?
4) In your opinion, what would be the most important aspects of the system, in regards to the personalisation of the feedback?
5) Could you see yourself using and benefitting from having access to an ATS? Why/why not?
6) In which areas do you think the ATS has an advantage/disadvantage over human tutoring?
7) Do you have any suggestions for changes or improvements of the ATS?

The following section will describe the findings from the conducted interviews, while highlighting some of the most noteworthy and relevant responses from the students.

Evaluation

In this section, the responses from the interviewed students are presented, and evaluated according to the three key areas mentioned above.

Experiences with and attitudes towards cognitive systems and personalised services

The interviews revealed that all students have had previous encounters with cognitive systems through mobile virtual assistants, such as Apple's Siri, Google Assistants, and Samsung's Bixby. There was a general consensus, that interacting with these systems left the students dissatisfied, and after the initial interactions, none of them use the service on a regular basis. This dissatisfaction was a result of different factors, such as; the language barrier; unnatural interactions; and limited functionality and capabilities of the virtual assistant, even with simple requests. Two of the students even stated the use was *"more of a gimmick, rather than a meaningful interaction"*. Moreover, it was mentioned that they lack specific use cases and clear purposes of interacting with the system, something which might motivate the students to use the system on a continuous basis in the future. Despite the rather striking criticism, there were very positive thoughts on the future of virtual assistants. It became clear that the students see a large potential and applicability for virtual assistants in the near future, and that it is a great way of applying and taking advantage of the technological capabilities in everyday life.

Viability, value creation, and potential use of the proposed ATS.

The preliminary interview in section 3.2 proved that the relationship between the student and teacher is somewhat professional and restricted. This has lead to the students in some cases refraining from asking questions about content or lectures that are not properly understood. In this interview, one student discussed the effect an ATS could have on this relationship: "*You act in a certain way [with a teacher, ed.] because you want to display knowledge and competence, and fear asking too many questions. By interaction with an ATS instead, such inhibitions are out of the way. You can approach the ATS more honestly, as you do not feel a constant judgement, or a need to prove yourself, as you would with a teacher. I see the ATS as a tool which can help solve this problem".* The remaining three students had similar attitudes towards how the ATS could assist, support, or even replace the student-teacher interaction. Based on their own experiences, two of the students noted that after finishing a lecture where the teacher has only touched upon certain topics superfluously, the accessibility of the ATS could help them find relevant answers to any questions that might arise. This highlights the importance of the domain model, which must be carefully implemented, to make the answers as concise and relevant as possible.

All of the interviewed students responded positively towards whether or not they could se themselves benefitting from having an ATS available. They did, however, also mention a number of factors which are crucial for the viability of the concept. One such factor, is the accuracy of the emotional analysis. *"The response must make sense, and not be pulled out of thin air",* one student said, and continued *"… the nonverbal signals must not be misinterpreted, nor should irrelevant signals should be considered."* An inaccurate analysis or erroneous processing of the emotional signals, would result in less appropriate answers from the ATS. Subsequently, the effect of the ATS will deteriorate. On a different note, the students also requested that the ATS must be accessible, and have a straightforward and natural interface, to make the interaction as seamless as possible. It was repeatedly suggested that the interface should be hands-free. Another factor is the quality of the response from the ATS. By quality it is understood that the response must correlate with the student's emotional state, be well-formulated, and presented in the same language as spoken by the student.

Besides the previously discussed application areas, the students indicated that they were most likely to use the ATS when preparing for a lecture, or studying for an exam. The ATS's immediate response to any surfacing question would be very welcome, especially when under certain time constraints, and when the need for an answer or feedback is urgent. One student suggested that analysing one's communication when preparing a presentation, would be another great use of the ATS, with the purpose of *"… getting feedback on your appearance, without being judged by teachers or fellow students.".* Another student saw the potential of gathering interaction data, and use it to evaluate or plan the courses, by looking for patterns across the questions being asked by different students.

Attitudes towards incorporating affect data and use of background information

Generally speaking, the students embraced the idea of taking nonverbal communication signals into consideration when determining the tutoring strategy and subsequent response. *"I like the idea of it [the ATS, ed.] reading my nonverbal signals, to decide whether I am in a mental state of frustration and need to be calmed down, or if I am more focused, and ready to be given an elaborate and thorough response to my question",* one student said. Another stated that making use of nonverbal signals *"…definitely makes sense, to construct more accurate and pedagogically correct responses (…) as made possible by the added affect information".* In regards to using physiological data, however, some concern was uttered throughout the interviews. One stated that *"if the extra information can give better feedback, then that is perfectly fine"*, but on the one condition that the data must be captured in an unobtrusive way, since "*extensive use of equipment would be disturbing, and definitely be a barrier of using the system".* Another student agreed with this statement, and said that *"the added value of physiological signals is not worth it, if it requires using a complicated setup with equipment and sensors".* The idea of using a camera or other "passive" monitoring equipment, which the student is not consciously aware of, was more positively received among the interviewees. Two students were sceptical of whether the cues sent when interacting with a computer, are actually comparable to their counterparts in interpersonal communication, e.g. with a teacher during class. One of them stated that his appearance is more neutral when interacting with a computer, compared to when he communicates in person, thus identifying a potential inaccuracy within the emotional readings.

The use of background information of the student, invoked some interesting discussions during the interview. First of all, all interviewees could see great potential in using this type of information when determining the tutoring strategy. One interviewee shared his thoughts on what kind of information he saw this module benefitting from: *"It could be everything from cold facts, such as age and gender, but also whether you are introvert or extrovert as a person, or whether you like to see things in writing, or as drawings and illustrations. (…) I think that this could yield extremely interesting results".* The aspect of incorporating communication and learning preferences, was also mentioned by a second student, who said that he was a "visual learner", so responses with videos or graphs would be a lot more valuable to him, compared to a regular written or spoken response. "*It could also be used in relation to how studying material is presented to the student in the first place, by incorporating individualised content, based on knowledge, learning style, and personality"*, a third student said. Furthermore, it was also added that a communication interface that uses the students mother tongue would be a great addition to the system.

In general, the students were open to incorporating personal information to a certain degree, but had huge concerns about privacy and protection of such personal data. One student stated that *"I am not afraid to give up personal information, as long as I benefit from it, and receive a better and more personal service",* however, the other interviewees where more reluctant towards the idea. *"Personal data should only be used if they are proven to greatly improve performance of the system",* one said. Another claimed that *"The more the system knows about the situational context and the student, the more interesting the interaction can be. But who has access to this information, and what else might the data be used for?".* These concerns were recurrent among two of the other students as well.

Key Findings

Based on the above mentioned responses, the following list shows the key findings from the interview:

- Despite previous experiences with virtual assistants, none of the students uses such services regularly, due to dissatisfaction with the language understanding, unnatural interactions, and limited capabilities of the systems.
- The use and application of virtual assistants have a high potential in the near future, but due to lack of use cases, the students have not yet adopted the technology.
- ATS could be a viable solution to bridge the gap within interpersonal communication between student and teacher, in regards to the students' inhibitions to ask questions.
- There is a positive attitude towards using ATS as an educational tool, but it must be ensured that the interaction is as seamless and natural as possible
- Accuracy of emotion signal processing is extremely important, and the relevance and quality of the response must be high, for the students to use the ATS on a continuous basis.

- The use of affect data from nonverbal signals was positively received, but it must not be at the expense of the user experience.
- There is a slight scepticism towards whether the same nonverbal signals are sent when interacting with a computer system compared to communicating interpersonally.
- Extensive background information could provide valuable information, especially in regards to individual learning styles and preferences.
- There is a concern about the privacy and data protection of personal data in regards to the student information module.

The learnings form these interviews showed that the general consensus among the interviewed students were that they definitely see a potential in using ATS in the future, if the systems are accurate, effective, and easy to use. Taking the possibilities of cognitive computing into consideration, the accuracy and effectiveness of systems such as ATS will most likely only improve from this point onwards.

6.5 Chapter summary

Based on the research conducted in chapters 2-5, a conceptual model of an Affective Tutoring Systems was presented. The model expanded upon the generic ATS architecture presented in section 3.1, by taking into consideration elements from the interpersonal communication process. The similarities and differences between this ATS model and the interpersonal communication model from section 4.8 were then discussed, according to the three aspects of understanding the student, interpreting the message, and responding appropriately. Furthermore, a second model was presented, illustrating and describing how different cognitive services could be used to implement a small-scale ATS.

Finally, follow-up interviews were conducted on the same four university students from the preliminary interviews, to evaluate the viability of the solution. This interview revealed that slight concerns about the accuracy of emotion analysis, user experience, and data protection, but that the general attitudes towards ATS were positive.

7. Discussion

From the preliminary research on intelligent and affective tutoring systems, and technological developments within cognitive computing, the following research question was formed:

> *How can cognitive computing technologies be used in Affective Tutoring Systems to emulate interpersonal communication between students and teachers?*

The findings proved that the interpersonal communication process to some extent can be emulated in an ATS system, by implementing cognitive services. While the methodological approach made it possible to answer the research question, it would be beneficial for the project to evaluate the project and proposed concept with experts in the field of cognitive and affective computing, or learning technologies such as ITS and ATS. Such evaluation could provide additional insights into areas such as: the technological developments of cognitive services, the potential of cognitive systems, insights into market tendencies of cognitive technologies, and the general business potential of ATS.

It can be argued that the proposed ATS model from section 6.1 has both strengths and weaknesses, when it comes to the the different components within the model. The starting point of the model was based on the generic architectural components of ATS, as presented by Malekzadeh, Mustafa and Lahsasna (2015) and Sarrafzadeh (2008), however, slight adaptations were made to conform with the interpersonal communication process. Since the focus of the project was mainly on incorporating affective information in a similar manner to interpersonal communication, the affect-related components were naturally discussed more in-depth compared to components less affected by emotions. Some of the components which were only briefly touched upon, were the emotion information processor module, the domain model, and the tutoring model. They are all essential components for the core functionality of the ATS, and must be further investigated to fully evaluate the potential of ATS. Determining the emotional state of the individual student is a crucial factor in determining how to respond, highlighting the importance of the emotion processor module. Some of the theories presented on human emotions can be used as a framework for processing the overall emotional state, where especially Russell's (1980) circumplex model of emotion is of interest. Similarly, the appropriate tutoring strategy should be incorporated to optimise the learning outcome, however, theory on learning processes and detailed pedagogical strategies is not within the scope of this project and has thus not been considered, but could be considered in future work.

The prototype model, describing the potential implementation of a system prototype, illustrated how cognitive services could be used in an ATS. While a system could be developed according to this model, this is by no means the only possible solution. Whereas the proposed model used facial expression analysis and emotion-mining from text to derive affect information, it could be interesting to explore the possibilities that comes with using several additional modalities. Additional speech capabilities, and physiological data could be incorporated with presumable large effects. Emotional analysis from speech is one of the areas in which the current technology is still far from human capabilities, as described in section 5.5. Emotion recognition in speech signals could be extremely interesting to incorporate into ATS, due to amount of affect information these signals carry in interpersonal communication. In the model, Microsoft's Emotion API is proposed as a facial expression analysis tool, and despite its impressive capabilities, more detailed emotional descriptions would be desirable. Furthermore, many ATS use animated agents to communicate responses and feedback to the student, and this is an aspect that could be interesting to explore as well.

The interviews in this project gave insights into the needs and attitudes towards ATS from the students' perspectives, however, it could be of great interest to approach the problem from the teachers' perspectives as well. This would provide additional insights into both the student-teacher relationship, and the need for ATS to support learning in both smaller and larger classes. Another way of expanding upon the empirical data gathered in the project, would be to interview students with different educational backgrounds and different educational levels. Another approach could be to gather additional quantitative data, which could support the qualitative findings from the interviews.

In the conducted interviews in section 6.4, the attitudes of students towards ATS were evaluated, however, this was based on a discussion of the conceptual models. To further support these interview findings, it could be beneficial to conduct an actual user test of an implemented ATS prototype. In the interviews, several students voiced their concern for the usability of a multimodal ATS, especially in relation to incorporation of physiological signals. Their concern was that galvanic skin response sensors or heartbeat sensors would be too obtrusive, and thus decrease the user experience of the ATS, which should be as natural and seamless as possible. The students were also sceptic towards whether or not the same signals are actually expressed in human-computer interaction as in interpersonal communication. A practical implementation would help address this issue as well. Another major concern which became evident from the interviews, was the issue of using personal data in the background information module. While the students saw a large potential in using this data to personalise feedback, they would also be reluctant to give access to such personal information, without a clear idea about who owns and access the data. This is an issue which must be addressed in a future implementation, as the personal data could provide extremely valuable information to both the emotion information processor module, and the tutoring model. Another solution to this issue would be to apply machine learning to interactional data from the ATS, which could enable profiling of the students based on their interactions instead of a background profile, thus reducing the need for comprehensive personal data, as such data might be inferred from student-specific pattern recognition algorithms instead. At this point, however, such an implementation is merely speculation.

8. Conclusion

Today, Intelligent Tutoring Systems (ITS), i.e. interactive systems capable of adapting to the individual learner, can be used to provide individualised tutoring. Using ITS has shown improved learning outcomes compared to most traditional forms of education, but still falls short of individual and small-group human tutoring, due to human teachers' ability to detect and respond to the affective state of the students. This has lead to the notion of Affective Tutoring Systems (ATS), which, as the name implies, also takes affect information about the student into consideration when deciding on a tutoring strategy for the individual. Background research was made within the fields of ITS and ATS to properly understand the capabilities of tutoring systems.

Interviews were conducted with four university students, which resulted in several interesting findings. First of all, it became evident that interpersonal communication with and feedback from teachers had a positive impact on the students' motivation and learning outcomes. Secondly, the needs, knowledge, and learning speed of the individual student is rarely considered, especially in larger classes. Finally, the students feel certain inhibitions related to the expectations from the teacher, and have a fear of demonstrating a lack of knowledge and exposing themselves, by asking questions to course material which has not been properly understood. The remainder of this project aimed to investigate how cognitive computing technologies could be used in an Affective Tutoring System (ATS) to emulate the interpersonal communication between students and teachers.

Based on communication theory and research on cognitive and affective computing, three models were created: one depicting the interpersonal communication process, another describing an ATS concept, and finally a model describing a small-scale ATS prototype. The first model (figure 14), illustrating the interpersonal communication process, was built based on extensive theoretical research within the fields of interpersonal communication and human emotions. The second model, the ATS (figure 16), extended the ATS architecture presented by Malekzadeh, Mustafa and Lahsasna (2015), by elaborating on the different components of the model, based on communication and emotion theory. Using the interpersonal communication model as a reference point, it was discussed to which extent the ATS concept emulated the interpersonal communication process, in regards to understanding the student, interpreting the student's message, and communicating an appropriate response. It became evident that many aspects of the communication process are also applicable in the ATS system, such as: considering individual background information when interpreting the student's messages; analysing both verbal and nonverbal communication signals in multiple channels; and adapting the response according to the detected emotional state of the student. That being said, many of the aspects and processes taking place in interpersonal communication are extremely difficult to represent, analyse, and replicate in a computer. Understanding and incorporating cultural context, apprehension, self-compliance, and message encoding, are examples of such complex processes.

The third model (figure 17) was created to illustrate how cognitive services could be used and combined in an ATS. By reviewing different cognitive development platforms, it became evident that a number of cognitive services were available for developers, which enabled e.g. emotion recognition from facial expressions, emotional tone analysis, and elaborate personality analysis. The functionality of these services are directly relatable to some of the affect recognition processes occurring in interpersonal communication.

Increased sets of training data, new technological developments, reduced hardware prices, and research and innovation within computer science, psychology, cognitive science and social sciences, are just some of the factors that will increase the capabilities of cognitive computing in the near future. Powered by deep learning and neural networks, cognitive systems have already proven capable of outperforming humans in a number of tasks that were previously unimaginable, thus making it difficult to argue why cognitive computing should not revolutionise the educational sector as well.

The findings from the project proved that it is possible to emulate the interpersonal communication process in an ATS system, to a certain degree. State of the art cognitive technologies enable processing of complex multimodal signals, similar to how communication signals are processed in the human brain, but their overall capabilities do not yet compare to the capabilities of the human brain. While processing power will only increase in the future, forecasting the upper limit of information processing capabilities within the system is extremely difficult.

9. Future work

Despite generating a number of interesting findings, there are still numerous areas which could be investigated, to generate further results and insights into the area of cognitive computing in ATS. The next step towards a more tangible project would be to implement the ATS proposed in figure 17. Implementing and evaluating an actual prototype would allow for more elaborate user testing, but it would also require additional knowledge on e.g. learning processes and pedagogical strategies, to proper implement the tutoring model of the ATS. It is also necessary to create a method or algorithm to handle the different API responses, and weigh and compute the emotional scores. This could be done by evaluating the emotions in an arousal-valence space, as presented in Russell (1980). Another relevant addition to the current project would be to include additional stakeholders, such as teachers, a more diverse group of students, or experts on cognitive technology and business development.

References

- Alexander, S., Sarrafzadeh, A., & Hill, S. (2008). Foundation of an affective tutoring system: Learning how human tutors adapt to student emotion. *International journal of intelligent systems technologies and applications*, 4(3-4), 355-367.
- Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. *Trames*, (2), 159-174.
- Andersen, P. A., & Guerrero, L. K. (Eds.). (1997). *Handbook of communication and emotion: Research, theory, applications, and contexts*. Academic Press.
- Argyle, M. (1975). *Bodily Communication*. New York: International Universities Press.
- Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., ... & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532-541.
- Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments*, 24(3), 590-605. DOI: 10.1080/10494820.2014.908927
- Bauer, C. (2014, October 27). *Beyond Siri, Google Now and Cortana: What is the future of virtual assistants?* Accessed 16/4/2017. Retrieved from https://thenextweb.com/insider/2014/10/27/beyond-siri-google-now-cortana-future-virtual-assistants/
- Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9(5), 691.
- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1), 1-127.
- Berlo, D. K. (1960). *The process of communication: An introduction to theory and practice*. New York, NY: Holt, Rinehart and Winston.
- Bochner, A. P. (1989). Interpersonal communication. In E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, & L. Gross (Eds.), *International encyclopedia of communications* (pp. 336-340). New York Oxford University Press.
- Chandler, D., & Munday, R. (2011a). *Verbal communication*. A dictionary of media and communication. OUP Oxford.
- Chandler, D., & Munday, R. (2011b). *Nonverbal communication*. A dictionary of media and communication. OUP Oxford.
- Cheney, G. (2011). *Organizational communication in an age of globalization: Issues, reflections, practices*. Long Grove, IL: Waveland Press.
- Clover, J. (2016, October 25). *Tim Cook: 'We Don't Buy' the Need to Give Up Privacy for AI.* Accessed 16/4/2017. Retrieved from https://www.macrumors.com/2016/10/25/apple-artificial-intelligence-privacy/

- Cognitive Computing Forum (2014). *Computing Forum 2014.* Accessed 16/4/2017. Retrieved from http://www.cognitivecomputingforum.com
- Corbyn, Z. (2016, January 31*). Meet Viv: the AI that wants to read your mind and run your life*. Accessed 16/4/2017. Retrieved from https://www.theguardian.com/technology/2016/jan/31/viv-artificial-intelligence-wants-to-run-your-life-siri-personal-assistants
- Cowie, R. (2005). *Emotion-oriented computing: State of the art and key challenges*. Humaine Network of Excellence.
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197-387.
- Earley, S., 2015. Cognitive computing, analytics and personalization. *IT Professional* 17(4), pp. 12-18. IEEE. DOI: 10.1109/MITP.2015.55
- Ekman, P. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system*. Salt Lake City, UT: A Human Face.
- Erez, A., & Isen, A. M. (2002). The Influence of positive affect on the components of expectancy motivation*. Journal of Applied Psychology*, 87(6), 1055–1067. doi:10.1037/0021-9010.87.6.1055.
- Equals 3 (2016). *How does Lucy work.* Accessed 30/4/2017. Retrieved from http://www.equals3.ai/howdoeslucywork
- Farías, D. I. H., Patti, V., & Rosso, P. (2016). Irony detection in Twitter: The role of affective content. *ACM Transactions on Internet Technology (TOIT)*, 16(3), 19.
- Foley, M. J. (2017, February 7). *Microsoft to make more Cognitive Services tools available to developers.* Accessed 6/6/2017. Retrieved from http://www.zdnet.com/article/microsoft-to-make-more-cognitive-services-tools-availabile-to-developers/
- Fujishin, R. (2008). *Creating communication: Exploring and expanding your fundamental communication skills*. Rowman & Littlefield Publishers.
- Gallois, C. (1993). The language and communication of emotion: universal, interpersonal, or intergroup?. *American Behavioral Scientist*, 36(3), 309-338.
- Gamble, T. K., & Gamble, M. W. (2013). *Interpersonal communication: Building connections together*. Sage Publications.
- Gangamohan, P., Kadiri, S. R., & Yegnanarayana, B. (2016). Analysis of Emotional Speech - A Review. *Toward Robotic Socially Believable Behaving Systems,* 1*,* 205-238. Springer International Publishing.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Google (2017). *Google Cloud Computing, Hosting Services & APIs*. Accessed 6/6/2017. Retrieved from https://cloud.google.com
- Graesser, A. C., Conley, M., & Olney, A. (2011). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 3. Applications to learning and teaching* (pp. 451–473). Washington, DC: American Psychological Association.

- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27-48.
- Hall, E. T. (1966). *The Hidden Dimension*. New York: Doubleday.
- Hassabis, D. (2016, March 16). *What we learned in Seoul with AlphaGo.* Accessed 30/4/2017. Retrieved from https://blog.google/topics/machine-learning/what-we-learned-in-seoul-with-alphago/
- Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion*, 6(3), 528-533.
- Hewlett Packard Enterprises (2017). *Haven OnDemand.* Accessed 16/4/2017. Retrieved from https://www.havenondemand.com
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554
- Hoey, J. (2017). *Lecture 2 - Overview of Emotions and Affective Computing.* Presentation, January 9, 2017. School of Computer Science. University of Waterloo.
- IBM (2014a, October 7). *IBM Watson: How it works*. Accessed 13/2/2017. Retrieved from https://www.youtube.com/watch?v=_Xcmh1LQB9I.
- IBM (2014b, February 24). *IBM Invests $1B to Deliver Unique Platform-as-a-Service Capabilities to Connect Enterprise Data and Applications to the Cloud*. Accessed 6/6/2017. Retrieved from http://www-03.ibm.com/press/us/en/pressrelease/43257.wss
- IBM (2017a). *Cognitive Computing.* Accessed 16/4/2017. Retrieved from http://www-03.ibm.com/software/products/en/category/SW77A
- IBM (2017b). *IBM Watson – Build your cognitive business with IBM.* Accessed 16/4/2017. Retrieved from https://www.ibm.com/watson/
- IBM (2017c). *IBM Watson Analytics*. Accessed 16/4/2017. Retrieved from https://www.ibm.com/analytics/watson-analytics/us-en/
- IBM (2017d). *IBM Watson for Oncology*. Accessed 13/2/2017. Retrieved from https://www.ibm.com/watson/health/oncology/
- IBM (2017e). *Catalog – IBM Bluemix*. Accessed 6/6/2017. Retrieved from https://console.ng.bluemix.net/catalog/
- IBM (2017f). *Tradeoff Analytics.* Accessed 23/3/2017. Retrieved from https://console.ng.bluemix.net/catalog/services/tradeoff-analytics/
- IBM (2017g). *About Speech to Text*. Accessed 28/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/speech-to-text/index.html
- IBM (2017h). *Overview of the Tone Analyser service*. Accessed 28/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/tone-analyzer/index.html
- IBM (2017i). *Understand your tone score*. Accessed 28/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/tone-analyzer/understand-tone.html
- IBM (2017j) *Personality Models*. Accessed 28/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/personality-insights/user-overview.html

- IBM (2017k). *Personality Models*. Accessed 28/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/personality-insights/models.html
- IBM (2017l). *Conversation | Getting started | IBM Developer Cloud.* Accessed 30/4/2017. Retreived from https://www.ibm.com/watson/developercloud/doc/conversation/getting-started.html
- IBM (2017m). *Text to Speech | About Text to Speech.* Accessed 30/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/text-to-speech/index.html
- IBM (2017n). *Text to Speech | The science behind the service.* Accessed 30/4/2017. Retrieved from https://www.ibm.com/watson/developercloud/doc/text-to-speech/science.html
- Irfan, M. T., & Gudivada, V. N. (2016). Cognitive Computing Applications in Education and Learning. *Handbook of Statistics*, 35, 283-300.
- Jackson, J. (2011, February 16). *IBM Watson Vanquishes Human Jeopardy Foes.* Accessed 16/4/2017. Retrieved from http://www.pcworld.com/article/219893/ibm_watson_vanquishes_human_jeopardy_foes.html
- Kelly, J. E., & Hamm, S. (2013). *Smart machines: IBM's Watson and the era of cognitive computing*. New York: Columbia Business School Publishing.
- Kendon, A. (2007). Some Topics in Gesture Studies. In A. Esposito (Ed.), *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, 18. IOS Press.
- Keyton, J. (2011). *Communication and organizational culture: A key to understanding work experience*. Thousand Oaks, CA: Sage.
- Kittlaus, D. (2016, May 9). *Beyond Siri: The World Premiere of Viv with Dag Kittlaus*. Accessed 16/4/2017. Retrieved from https://www.youtube.com/watch?v=Rblb3sptgpQ
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42-78.
- Langton N., & Robbins S. P. (2006). *Fundamentals of Organizational Behavior*. Ontario: Pearson.
- Lane, S. D. (2008). *Interpersonal communication: Competence and contexts.* Pearson.
- Langlotz, A., & Locher, M. A. (2013). The role of emotions in relational work. *Journal of Pragmatics*, 58, 87-107.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2), 341-348. doi:10.1016/j.mehy.2011.11.016. PMID 22153577.
- Lunenburg, F. C. (2010). Communication: The process, barriers, and improving effectiveness. *Schooling*, 1(1), 1-11.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology* 106(4), 901-918. American Psychological Association

- Maiocchi, M. (2015). Perception and Emotions. *The Neuroscientific Basis of Successful Design*, 25-40. Springer International Publishing.
- Malekzadeh, M., Mustafa, M. B., & Lahsasna, A. (2015). A Review of Emotion Regulation in Intelligent Tutoring Systems. *Educational Technology & Society*, 18(4), 435-445.
- Manning, C. & Socher, R. (2017). *CS224n: Natural Language Processing with Deep Learning*. Stanford University.
- Martin, E. A. (2010a). neuron. In *Concise Medical Dictionary*, Eighth edition.
- Martin, E. A. (2010b). synapse. In *Concise Medical Dictionary*, Eighth edition.
- Merriam-Webster (2017). *Cognition*. Accessed 16/4/2017. Retrieved from https://www.merriam-webster.com/dictionary/cognition
- Microsoft (2016). *Cognitive services APIs.* Accessed 16/4/2017. Retrieved from https://www.microsoft.com/cognitive-services/en-us/
- Microsoft (2017a). *Cognitive Services.* Accessed 6/6/2017. Retrieved from *https://azure.microsoft.com/da-dk/services/cognitive-services/*
- Microsoft (2017b). *Microsoft Cognitive services - Emotion API.* Accessed 28/4/2017. Retrieved from https://www.microsoft.com/cognitive-services/en-us/emotion-api
- Miell, D. E. (1984). *Cognitive and communicative strategies in developing relationships.* Doctoral dissertation, University of Lancaster.
- Mongrain, M., & Vettese, L. C. (2003). Conflict over emotional expression: Implications for interpersonal communication. *Personality and Social Psychology Bulletin*, 29(4), 545-555.
- Motahari, H. (2014, March 19). *Cognitive Work Assistants - Vision and Open Challenges.* [PowerPoint Presentation]. Accessed 16/4/2017. Retrieved from https://www.slideshare.net/hrmn/cognitive-assistants-vision-and-open-challenges
- Nesbitt, K., Blackmore, K., Hookham, G., Kay-Lambkin, F., & Walla, P. (2015). Using the startle eye-blink to measure affect in players. In *Serious Games Analytics*, 401-434. Springer International Publishing
- Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy. *Journal of Nonverbal Behavior*, 18, 9–35.
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the influence of thought and of the science of symbolism*. New York: Harcourt, Brace & World, Inc.
- Oxford Dictionaries (2017). *Cognition*. Accessed 16/4/2017. Retrieved from https://en.oxforddictionaries.com/definition/cognition
- Pantic, M., & Rothkrantz, L. J. M. (2003). Toward an Affect-Sensitive Multimodal Human-Computer Interaction. *Proceedings of the IEEE, 91*(9). IEEE.
- Parrott, W. (2001). Emotions in Social Psychology. *Key Readings in Social Psychology*. Philadelphia: Psychology Press. ISBN 978-0863776830.
- Partan, S. R., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist*, 166(2), 231-245.
- Pereira, D. T., Schatsky, D., Sallomi, P. & Dalton, R. (2015, December 15). *Cognitive technologies in the technology sector. From science fiction vision to real-world value*. Accessed

16/4/2017. Retrieved from https://dupress.deloitte.com/dup-us-en/focus/cognitive-technologies/artificial-intelligence-in-technology-sector-tmt.html

- Pesenti, J. (2014, December 4). *Cognitive Computing.* Accessed 6/6/2017. Retrieved from https://www.youtube.com/watch?v=8zYp4yH4PoQ
- Pipas, M. D., & Jaradat, M. (2014). Perception of emotion in interpersonal communication. *Calitatea*, 15(S3), 96-101.
- Planalp, S. (1998). Communicating emotion in everyday life: Cues, channels, and processes. *Handbook of communication and emotion: Research, theory, applications, and contexts*, 29-48.
- Plutchik, Robert (1980), Emotion: Theory, research, and experience: Vol. 1. *Theories of emotion*, 1, New York: Academic.
- Protalinski, E. (2017). *Google's speech recognition technology now has a 4.9% word error rate*. Accessed 30/5/2017. Retrieved from https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/
- Qadar, F. (2016). *Self-Talk: Effects on Emotion in Interpersonal Communication Context*. The University of Arizona.
- Raghavan et al., (2016). *Cognitive Computing: Theory and applications*. Elsevier.
- Roe, C. (2014, May 8). *A Brief History of Cognitive Computing*. Accessed 16/4/2017. Retrieved from http://www.dataversity.net/brief-history-cognitive-computing/
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). Sensitivity to nonverbal communication: The PONS test. Baltimore: John Hopkins University Press.
- Rouse, M. (2016). *Cognitive Computing.* Accessed 16/4/2017. Retrieved from http://whatis.techtarget.com/definition/cognitive-computing
- Rubin, R. B., Perse, E. M., & Barbato, C. A. (1988). Conceptualization and measurement of interpersonal communication motives. *Human Communication Research*, 14(4), 602-628.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. doi:10.1037/h0077714
- Samsung (2016, October 6). *Samsung to Acquire Viv, the Next Generation Artificial Intelligence Platform.* Accessed 16/4/2017. Retrieved from https://news.samsung.com/global/samsung-to-acquire-viv-the-next-generation-artificial-intelligence-platform.
- Saon, G. (2017, March). *Reaching new records in speech recognition.* Accessed 30/5/2017. Retrieved from https://www.ibm.com/blogs/watson/2017/03/reaching-new-records-in-speech-recognition/
- Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *Computers in Human Behavior*, 24(4), 1342-1363.
- Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns?. *Emotion*, 7(1), 158.

- Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index. *Journal of Nonverbal Behavior*, 35(4), 305.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- Schramm, W. (1954). How communication works. In W. Schramm (Ed.), *The process and effects of communication*, 3-26. Urbana, Illinois: University of Illinois Press.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.
- SoftBank Robotics (2017). *Pepper robot: characteristics.* Accessed 30/4/2017. Retrieved from https://www.ald.softbankrobotics.com/en/cool-robots/pepper/find-out-more-about-pepper
- Sottilare, R., Graesser, A., Hu, X., & Holden, H. (Eds.). (2013). *Design recommendations for Intelligent Tutoring Systems*. Orlando, FL: U.S. Army Research Laboratory.
- Stamp, G. H. (1999). A Qualitatively Constructed Interpersonal Communication Model: A Grounded Theory Analysis. *Human Communication Research*, 25(4), 531-547.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of Intelligent Tutoring Systems (ITS) on college students' academic learning. *Journal of Educational Psychology*, 106, 331–347. doi:10.1037/a0034752
- Sudhkar, R. S., & Anil, M. C. (2016). Emotion Detection of Speech Signals with Analysis of Salient Aspect Pitch Contour. *Emotion*, 3(10).
- Sun X., Wu, P., & Hoi, S. C. (2017). *Face Detection using Deep Learning: An Improved Faster RCNN Approach*. Accessed 8/6/2017. Retrieved from https://arxiv.org/pdf/1701.08289.pdf
- Tannen, D. (1986). *That's not what I meant!*. New York: Ballantine.
- Tao, J., & Tan, T. (2005). Affective Computing: A Review. *Affective Computing and Intelligent Interaction Lecture Notes in Computer Science*, 981-995. doi:10.1007/11573548_125
- Tung, L. (2016, September). *Microsoft's newest milestone? World's lowest error rate in speech recognition.* Accessed 30/5/2017. Retrieved from http://www.zdnet.com/article/microsofts-newest-milestone-worlds-lowest-error-rate-in-speech-recognition/
- VanLehn, K. (2011). The relative effectiveness of human tutoring, Intelligent Tutoring Systems, and other tutoring systems. *Educational Psychologist*, 46, 197–221. doi:10.1080/00461520.2011.611369
- Viola, P., & Jones, M. (2001). Robust real-time face detection. *Proceedings Eighth IEEE International Conference on Computer Vision*. ICCV 2001. doi:10.1109/iccv.2001.937709
- Viv (2017). *Viv.* Accessed 16/4/2017. Retrieved from http://viv.ai
- Waters, R. (2016, September 26). *Siri, Alexa, Cortana and the unstoppable rise of the digital assistant*. Accessed 16/4/2017. Retrieved from http://www.afr.com/technology/apps/business/siri-alexa-cortana-and-the-unstoppable-rise-of-the-digital-assistant-20160925-grnxvj
- Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Computing Surveys (CSUR), 50*(2), 25.