UNIVERSITY OF TWENTE.

Nonverbal Behaviour of an Embodied Storyteller

F.Jonkman
f.jonkman@student.utwente.nl

Supervisors:
Dr. M. Theune, University of Twente, NL
Dr. Ir. D. Reidsma, University of Twente, NL
Dr. D.K.J. Heylen, University of Twente, NL

ABSTRACT

A system that is visualized as a person and has some communicative capabilities is often called an Embodied Conversational Agent. Some of these Embodied Conversational Agents show a limitation in their available gaze behaviours towards the user. Take for example agents that are positioned directly in front of the user. Most of them are limited to a look-at the interlocutor and look-away from the interlocutor state as their available gaze behaviours. The question arises if and how these behaviours of the Embodied Conversational Agents can be improved and be extended with other gaze behaviours? Will the introduction of other gaze states improve the viewing experience? In this master thesis two viewing perspectives and the accompanying gaze behaviours are researched and the question is asked whether there is a preference for one of the two viewing perspectives.

The first perspective is the newsreader perspective, often found in existing research aimed at analyzing the verbal and non-behaviours of an Embodied Conversational Agents. The newsreader perspective is a viewing perspective in which the embodied agent is located directly in front of the camera. The second viewing perspective is the storyteller perspective, seen in TV-series like Sesame Street and the Dutch TV-series "Elly en de wiebelwagen". The viewing perspective introduces an audience located directly in front of the storyteller with a sideways camera placement. Subsequently gaze-behaviours are introduced as the main non-verbal behaviours of the embodied storyteller. By analyzing existing gaze models and annotating an actual storyteller a gaze model is defined for the embodied storyteller. The gaze model and the viewing perspectives were implemented in the verbal and non-verbal behaviour realizer Elckerlyc [WR10].

The gaze behaviours and the viewing perspectives are evaluated, by means of a short user-survey. The results of the survey show a slight preference for the storyteller perspective and none of the gaze behaviours are valued as purely negative. However the most important thing the user-survey shows is the large variance in participant's opinions about suitable gaze behaviours and preference for one of the two viewing perspectives. The question arises if there is such a thing as an "perfect" viewing experience with suitable non-verbal behaviours from a viewing perspective, that meets the expectations and preferences of most users.

PREFACE

The project regarding the available non-verbal behaviours of the Virtual Storyteller [ST08] grabbed my attention, because it provided me with the possibility to research human-like gaze behaviours and to recreate them in a virtual environment. Previous experience in creating a turn-taking model for agent-dialog in the Virtual Storyteller and the influence of head-movements in Social Signal Processing provided me with some background information about non-verbal behaviours, their function and meaning.

User's behaviours and expectations towards new types of applications are for me one of the most captivating and hard to understand subjects. Experience from mobile app design at my current employer and especially the knowledge gained during the evaluation phase, taught me how difficult it is to design something that is valued by everyone.

I started the master-thesis with the clear expectation that there would be a noticeable preference for the storyteller perspective. However in the end the user-survey provided me with the understanding on how diverse user's opinions are. Not only regarding preference for a certain viewing perspective, but also the embodiment of the storyteller and the expectations participants have with certain gaze-behaviours.

LIST OF CONTENTS

1. INTRODUCTION

1.1 Background

For quite some time now researchers are trying to model human behaviour displayed during a conversation or interaction. These behaviours, either verbal or nonverbal, can be used by a computer system to augment the interaction it has with the user. Current advances in computational and graphical processing power provide the computer system with the capabilities to look and act more human-like every day. Depending on the situation these behaviours can be used to present information to an audience, consult a client or tutor a student. Take for example a (nonhuman) tour guide that is provided at the museum. Early tour guides where simple booklets that provided some basic information at certain points in the tour. Next there were the audio tapes with a storyteller that guided you during your tour. Nowadays we have the ability to create a virtual representation of an human tour guide, that not only looks like an actual person but also acts like it and is fully interactive.

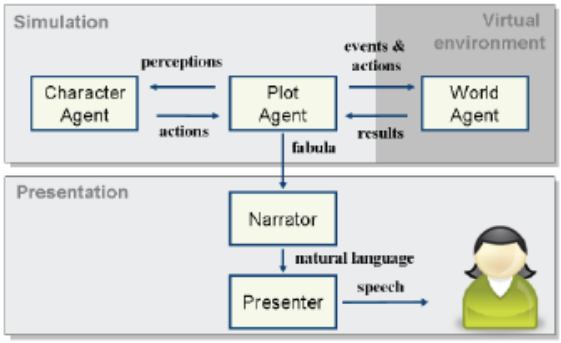

Figure 1: Global architecture of the Virtual Storyteller [ST08]

Systems that represent humans and that model the accompanying human-like behaviours are often called Embodied Conversational Agents [JC01]. They have the ability to use one or more of the modalities humans use during a conversation. These modalities can be for example the variations in speech, facial expressions, gaze behaviours, head movements, hand gestures, etc. An example of an embodied agent that models verbal and nonverbal behaviours, is the presenter agent from the Virtual Storyteller [ST08]. The Virtual Storyteller is a multi-agent framework (figure 1), that uses characters in a virtual environment to generate stories based on simulation. The sequence of events, the current state of the different characters and information about the environment are all used to produce a story. The narrator agent uses this information from the simulation layer, in combination with its knowledge about language generation techniques to generate a story. The presenter agent has the possibility to use synthesized speech to tell the generated story to an audience. It can use certain facial expressions to enhance the experience. Determining the correct facial expressions is done by tagging the story with the specific moments when appropriate facial expressions can be performed.

To further improve the nonverbal behaviours of the presenter agent, other modalities should be considered to improve the storytelling of the agent. These modalities can be for example the use of correct gaze behaviours, head movements or correct hand gestures. Not only do humans use these nonverbal behaviours in their daily interactions, humans expect certain nonverbal behaviours during a conversation. Although the behaviours are not consciously processed, they are an integral part of creating a pleasant viewing experience and make the story more memorable. The presenter agent should use these behaviours to capture the attention of the audience and keep them engaged in the story. If the audience loses the interest in the story, the presenter agent loses the attention of the audience and subsequently the content of the story is lost. Correct use nonverbal behaviours can make or break the experience.

For this master thesis the gaze behaviours of a real storyteller will be investigated. The determination and use of correct gaze behaviours will be researched and will be used by an embodied agent that represents an actual storyteller. Besides the gaze behaviours, the perspective from which the story is viewed will be part of the analysis. The presenter agent from the Virtual Storyteller talks directly towards the camera (figure 2), the hypothesis is that perhaps this perspective is not the most interesting perspective from which the story can be viewed. From daily TV-shows (Sesame Street) it seen that there are viewing perspectives in which an audience is present and there is a sideways viewing perspective for the user (figure 3). From these TV-shows it is also seen that the story can be read from a book. The perspectives combined with the appropriate gaze behaviours, provides a situation with different elements that influence the viewing experience for the user in one way or the other.

1.2 Research Question

From the information provided in the previous section a main research question has been derived. The main research question provides the starting point for the master-thesis.

When creating viewing experiences from two different perspectives, in which the gaze behaviour is the main nonverbal behaviour, which of the perspectives provides the best viewing experience for storytelling?

The two viewing perspectives are the newsreader perspective (figure 2) and the storyteller perspective (figure 3). Where the first is a frontal viewing perspective (as found in the presenter agent from the Virtual Storyteller) and the latter is a sideways viewing perspective (as found in different TV-shows). From the main research question multiple sub-questions can be derived:

- What are the most important gaze behaviours displayed by an actual storyteller during the telling of a story?
- When recreating the storytelling situation in a virtual environment, which elements need to be modelled and how? (The storyteller, the gaze behaviours, the book, the audience, etc.)
- How do users valuate both viewing perspectives and the accompanying gaze behaviours?

Figure 2: The newsreader perspective

Figure 3: The storyteller perspective

1.3 Method

The main objective for this master thesis will be to derive the possible gaze behaviours of an embodied storyteller during storytelling and examine two possible viewing perspectives. The perspectives consist out of a newsreader and storyteller perspective. It is not the objective to "copy" the exact gaze behaviour of the real storyteller, but to derive the main characteristics of the displayed gaze behaviours (i.e. the duration of the gaze behaviours and their gaze frequencies). These characteristics will be tested in a virtual environment by creating a gaze model for the gaze behaviours.

The first step will be to investigate the current state of existing Embodied Conversational Agents. The most important elements will be written down. Existing literature in nonverbal behaviours as a whole and its implications on a conversation will be researched. Subsequently existing gaze models will be reviewed. The existing gaze models will serve as a starting point for our own gaze model. The gaze model will be used by an embodied storyteller in our own storytelling situation, in which the different viewing perspectives will be assessed.

Also the gaze behaviours of an actual storyteller will be determined by annotating existing video material, in which an actual storyteller tells a tale towards an audience and the user. From the literature and the annotated sessions the most important gaze behaviours will be specified. The gaze frequencies and the gaze durations of the different gaze behaviours from the annotations will be used in the creation of our own gaze model. The embodied storyteller will use this gaze model during the telling of story to create its own gaze behaviours.

To test both viewing perspectives an storytelling situation will be recreated in a virtual environment. This will be done by using Elckerlyc [WR10], a verbal and non-verbal behaviour realizer. Elckerlyc uses the Behaviour Mark-up Language (BML) to simulate multiple nonverbal behaviours, including the gaze behaviours. As mentioned before two viewing perspectives will be created:

- The newsreader perspective: the embodied storyteller tells the story from a book, talks directly towards the camera (user) and uses gaze behaviours as its main nonverbal behaviour.
- Storyteller perspective: the embodied storyteller tells the story from a book, talks directly towards an audience, the camera (user) is placed at a sideways position next to the audience and gaze behaviours are its main nonverbal behaviour.

To review both viewing perspectives and assess the displayed gaze behaviours performed by the embodied storyteller, a user survey will created. The user survey which will address the following questions:

- Does the gaze model provide gaze behaviours, that meet the user's expectations of an actual storyteller?
- How are the different elements forming the viewing experience rated? (the gaze behaviours, the introduction of an audience, the viewing perspective, the embodied storyteller itself, etc)
- How does an embodied storyteller, that tells the story directly towards a camera, relate to an embodied storyteller, that tells the story towards an audience? Does the latter viewing perspective provide a better viewing experience?

1.4 Thesis Overview

The thesis will start with a general definition of an Embodied Conversational Agent and two existing embodied storytellers are mentioned. Next the question will be answered if users perceive Embodied Conversational Agents as real conversational partners or as purely information providing interfaces. Subsequently there is a description of the nonverbal behaviours and two existing gaze models will be used to describe gaze behaviours in more detail. Subsequently an actual storyteller and its gaze behaviours during storytelling is annotated. The data collected during these annotations is used in the formation of our own gaze model. The gaze model than implemented in a Virtual Human in an existing framework that makes the realization of verbal and nonverbal behaviours possible. The evaluation of the viewing perspectives, the gaze behaviours and the viewing experience is done by performing a user survey and the results will be analysed in detail. The thesis will end with a conclusion and some possible future work.

2. EMBODIED CONVERSATIONAL AGENTS

"Embodied Conversational Agents - an interface in which the system is represented as a person, information is conveyed to human users by multiple modalities such as voice and hand gestures, and the internal representation is modality independent and both propositional and non-propositional [CJ01, p67]"

There are multiple examples of existing Embodied Conversational Agents. Early Embodied Conversational Agents include *REA* developed by Cassell et al. [CJ01] and *Gandalf, the Interactive Guide to the Solar System* developed by Thorisson [TK02]. However because the focus of this master thesis is on the non-verbal behaviour displayed during storytelling and not during a conversation, two existing embodied storytellers will be discussed. Although both lack a clear description regarding the selection of the verbal and non-verbal behaviours, they do provide some background information in the current state of embodied storytellers.

2.1 Embodied Storytellers

Storytellers tell stories and use their verbal and non-verbal behaviours to keep the attention of the audience. We will begin with discussing two existing Embodied Conversational Agents that represent storytellers and both use different verbal and non-verbal behaviours to augment the experience. The first embodied agent is Papous: the Virtual Storyteller and should come close to a storyteller we want to create in a virtual environment. The next agent is the storyteller SAM that tries to aid children during the creation of a story.

Papous: The Virtual Storyteller

"A good storyteller is able to drag us into the story, keep our attention and free our imagination [SV01, p171]"

Papous: The Virtual Storyteller is an example of an embodied storyteller that has the ability to perform different types of non-verbal behaviours during the telling of a story, to make it more expressive and believable. Silva et al. [SV01] summarized the different aspects of how a good storyteller can transform a simple story into an enriched storytelling experience. They mention the use of the voice, facial expressions and the appropriate gestures as the main ingredients that can transform a simple story into a interesting narrative.

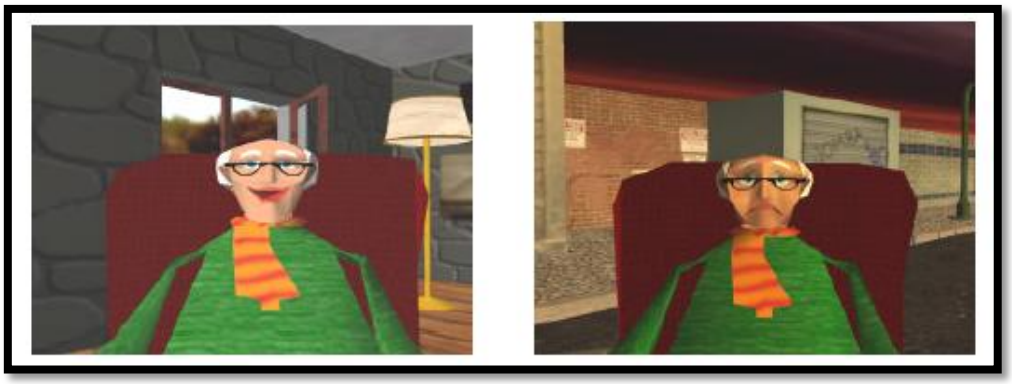

Figure 4: Papous who is happy (left) or sad (right) [SV01, p177]

These three ingredients (user of the voice, facial expressions, gestures) have been implemented into Papous and during the telling of the story Papous reads text augmented with different tags that control its behaviour, the scene in which the story takes place, the illumination of the scene and the current emotion of the virtual agent (figure 4 and figure 5). There is no lip synchronization, only a gesture that opens en closes the mouth sequentially, also the TTS System does not provide great deal of flexibility during the telling of a story.

Papous serves as an example of an embodied storyteller that uses different non-behaviours to improve the storytelling experience. One would expect that this embodied agent comes close an embodied storyteller that will created in this master thesis. However Silva et al. never fully explain the selected gestures, emotions and facial expressions during the telling of a story. It lacks argumentation and background information on the importance of the selected facial expressions and hand gestures. Why are they the determining factors for the experience?

Figure 5: Papous indication something big (left) and something small (right) [SV01, p178]

SAM

The next example of an embodied storyteller is SAM which tells stories to children interactively [RV02]. SAM was developed to aid children with literacy learning and has the ability to listen to children's stories. SAM is an embodied storyteller projected on a screen behind a castle (figure 5). SAM starts by welcoming and looking at the child, next SAM tells a story whilst moving a figurine (which exists both in the virtual and physical world) through the castle. When the SAM finished with her story, she asks the child to open a door in the castle. Behind this door the figurine is located and the child is asked if he or she can tell a story. During the this time SAM watches the child and is nodding, smiling and prompting "What happens next?". When the child is done the figurine is given back to SAM and the interaction continues. So the gaze of SAM is at the child itself and the object the child is moving through the castle and is made possible by using microphones, motion sensor detectors and RFID tag readers to determine the location of the figurine. However a clear description of the used gaze model is absent.

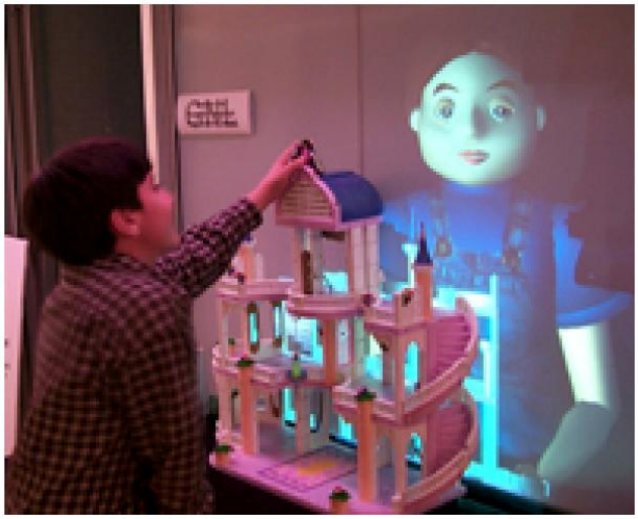

Figure 6: SAM [CJ01]

Through different experiments with the children it is seen that SAM is accepted as an interactive partner during the creation of their own story. The children's stories became more sophisticated and complex by what they had learned from the stories told by SAM. Children not only regarded SAM as a conversational partner, but saw SAM as a person who needed to be aided during the storytelling. If the story told by SAM was too short, the children would comment to make the story longer next time. Yet from the results it isn't apparent which of the verbal or non-verbal elements are most important for the acceptation of SAM as a conversational partner.

2.2 Conversational Partners

If humans talk to each other, they use more than their speech to present the information. The interaction between the interlocutors is a complex process consisting of multiple modalities, that are mainly processed or subconsciously. Humans use for example their eye gaze to transmit speaker and listener turns (interactional function) and utilize head movements to indicate agreement or disagreement (propositional function). Nonverbal behaviours are not consciously retained, but add to the whole of the experience. Absence of the nonverbal behaviours during the interaction between an embodied agent and the user, results in users starting to repeat themselves more often and judge the systems use of language and the understanding of language as worse [CJ01].

Louwerse et al. [LG08] have reviewed several studies which researched the interaction between Embodied Conversational Agents and their users. Their main focus was on how humans think and interact with an Embodied Conversational Agent. The results from the different studies were very diverse. On the one hand children benefit from the interaction with an embodied conversational agent over text,- or voice-only learning, on the other hand there are studies that show a less conclusive picture of the usefulness of the introduction of an embodied conversational agent. One study showed that the embodied conversational agent had no added effect over print-alone or speech-alone learning [LG08]. Another study showed that there is little to no difference between the results of using a fully dynamic or a minimally static embodied conversational agent.

Louwerse et al. asked themselves the question how users that interact with an embodied conversational agent pay attention to the embodied conversational agent and provided two hypotheses:

1. The information-only hypothesis: Embodied Conversational Agents will not attract or will lose attracted attention if they do not provide additional information to the communication.
2. The conversational-partner hypothesis: users look for perceptual cues throughout the interaction with an embodied conversational agent to guide their attention.

They tested both hypotheses by tracking the eye movements of different users during the interaction with two different Embodied Conversational Agents. The first embodied conversational agent was AutoTutor (figure 7), an intelligent tutoring system that educates students by holding natural language interactions.

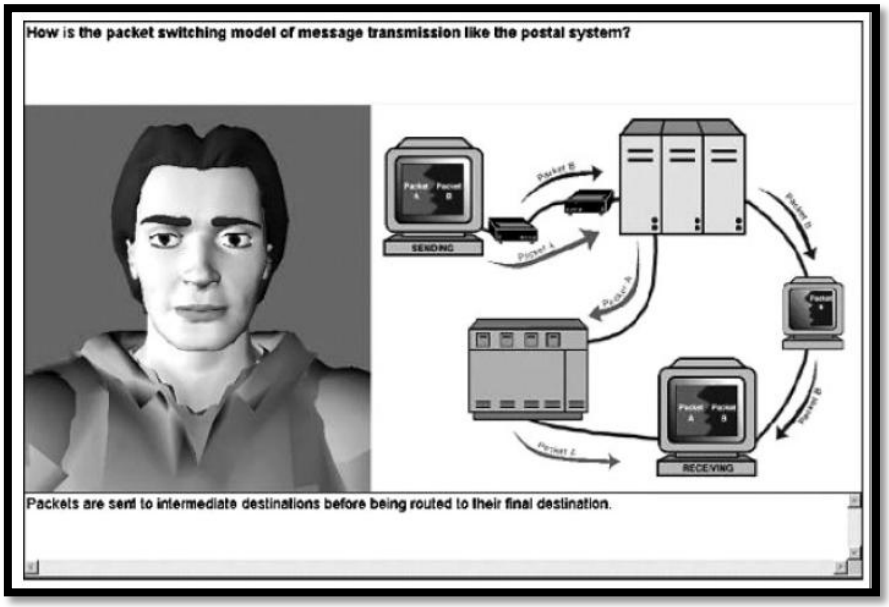

Figure 7: AutoTutor interface with four information sources: ECA, question, student input and graphic display [LG08]

From the information-only hypothesis it is to be expected that user would only listen to the embodied conversational agent and would focus on the other information resources. According to the conversational-agent hypothesis users would look at the embodied conversational agent because it serves as a participant in the dialog. The results showed that the users regarded the embodied agent as a conversational partner, because they often looked at the agent during the conversation.

In the second study they tested a group of Embodied Conversational Agents and the prediction was that the eye gaze of the users will be on those Embodied Conversational Agents relevant to the conversation. They used the intelligent tutoring system iSTART (figure 8). This system helps students to self explain texts while reading. During the introduction phase of the system there are three Embodied Conversational Agents interacting with each other and the user. There are two components that capture the attention of the user, namely the three characters in the screen and a text balloon above the characters as soon as they start to speak. The second study investigated if users looked at the relevant agent at the correct time, supporting the conversational-partner hypothesis. On the other hand if users would not look at the relevant agent at the correct time, would support the information-only hypothesis. The second experiment also showed that the Embodied Conversational Agents have the effects of a conversational partner and that during the entire interaction the user will look at the correct embodied conversational agent.

Figure 8: iStart interface with three agents

The studies suggest that users see the Embodied Conversational Agents as conversational partners. This means that from a social perspective the users have certain social expectations about how the conversation will proceed. Figure 9 shows a number of the nonverbal cues humans transmit during a conversation, even with the absence of the audio is it clear that these two humans are in heavy conversation. To make computers socially intelligent, they must have an understanding about the nonverbal behaviours humans use in their interactions. The nonverbal behaviours a human display during a conversation are mostly honest, reliable and above all done unconsciously. Subsequently they provide the embodied conversational agent not only with a lot of useful feedback, but the agent can use knowledge about the influence of presented nonverbal behaviour to its own benefits by controlling its own nonverbal behaviour.

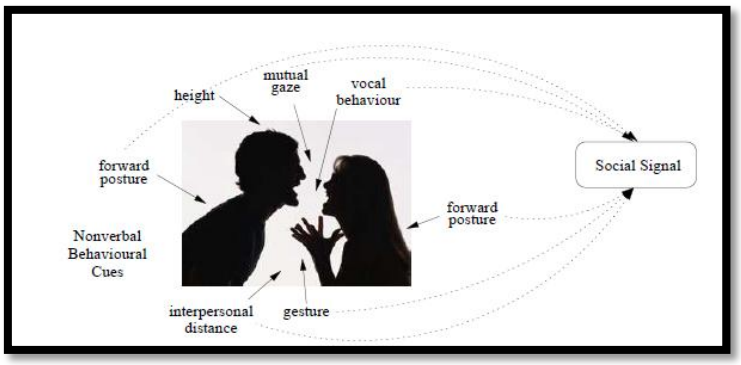

Figure 9: Behavioural cues and the social signals [VP08]

The two existing Embodied Storyteller, Papous [SV01] and SAM [RV02], serve as a small introduction to the field of Embodied Agents used as actual storytellers. The paper of Louwerse et al. [LG08] demonstrate that users perceive the Embodied Conversational Agents as conversational partners and that there are social expectations that come with the interaction. So the Embodied Storyteller should be perceived as an actual storyteller and its gaze behaviours should reflect the gaze behaviours of an actual storyteller.

3. NONVERBAL BEHAVIOUR

Nonverbal behaviours are an integral part of the communication process. They are predominantly honest, reliable and done unconscious. They provide the user with information that can be useful to the interaction and form a basis for a social relationship between the interlocutors. Vinciarelli et al. [VP08] describe an upcoming research field that tries to create intelligent systems that not only have the ability to capture the non-verbal behaviours someone displays, but also have the ability to understand the social signal that is being transmitted (Social Signal Processing).

> *In the case of social interactions, nonverbal communication takes the form of social signals [2][3], complex aggregates of behavioural cues accounting for our attitudes towards other human (and virtual) participants in the current social context.* [VP08, p1]

Although Social Signal Processing and the understanding of a nonverbal behaviour for the embodied agent is beyond the scope of this master thesis, the provided description of the different nonverbal behaviours with their cues, their code and their function is worth mentioning. Some of the nonverbal behaviours will be used by the embodied agent and it is useful to know to what end humans use them in their daily interactions (figure 10).

Nonverbal behavioural cues: Visible changes in facial expressions and body gestures that accompany our communication with each other. Codes: These are groups of nonverbal behaviours that have the same function. Functions: The specific function for a code of nonverbal behaviours.

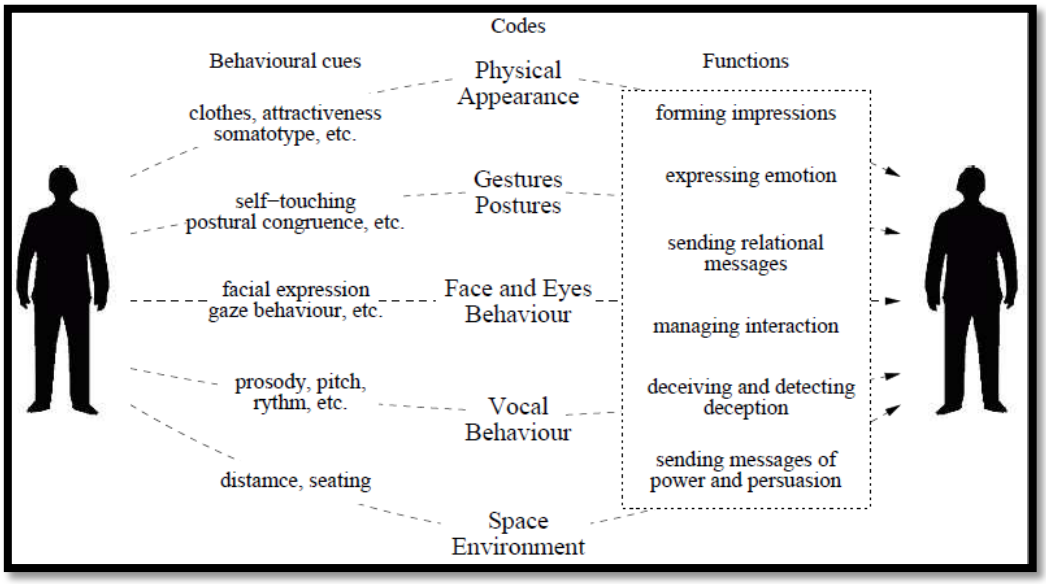

Figure 10: Behavioural cues, codes and functions. Nonverbal behavioural cues are organized into codes and fulfil functions aimed at affecting the perception of others [VP08, p3]

For this master thesis we are interested in the different gaze behaviours a human displays during a social interaction. The gaze behaviours belong to the code of "Face and eyes behaviours". Examples of the function of different gaze behaviours are:

- Forming impressions: gaze away shows insecurity.
- Used to manage the interaction: gaze towards the listener to release speaker turn.
- Express emotions: gaze downwards when sad.
- Send relational messages: social conventions like nodding and gazing, during a conversation.
- Deceive and detect deception: fast gazing when lying.
- Send messages of power and persuasion: powerful people gaze less towards others.

The next section will take a closer look at head movements and they are used during a gaze behaviour. Subsequently two exiting gaze models will be discussed. The gaze models show how gaze behaviours can be constructed and how variances in the gaze behaviour of an agent influences the valuation of the user of the agent.

3.1 Head Movements

Head movements provide a rich source of conversational feedback. Where the nods and shakes are the most common head movements, accompanied by tilting of the head. Nods can signal agreement, understanding, approval, etc. Shakes can mean disagreements, misunderstanding, disapproval, etc. The tilting of the head can be divided in to a sideways tilt, which can indicate submissiveness or friendliness, an upwards tilt which can indicate arrogance or superiority [GD02] and finally a downwards tilt, which can indicate insecurity or shame. Besides the movements and the meaning of the head movements there is the function of the head movement. Heylen [HD05] describes three main categories:

- There are head movements important for the information structure
- There are head movements that provide expressive reactions
- There are head movements important for the interactional management (turn taking).

Heylen [HD05] further discusses research in which head movements are linked to speech. It states that head movements mostly occur when the person is speaking and that during listener turns the head is mostly still. Rapid head movements have been linked to primary peaks of loudness of speech. The louder someone talks the faster the head movements will occur. Not only the speaker's head movements have been linked to the speaker's utterances, but also the listener's head movements are linked to the utterance of the speaker. Individual head movements can have multiple explanations and this makes it difficult to determine the appropriate head movement at a certain time without the availability of certain context. Head movements require a certain understanding of the story and the current setting in which the story is told.

For this master thesis we are interested in the head movements that can be context free. These are the head movements performed during a gaze behaviour. In a gaze behaviour not only the eyes move, but also the entire head can shift towards the selected gaze point. That gaze behaviours do not necessarily require context about the story will be discussed in the upcoming section by reviewing two existing gaze models.

3.2 Gaze Behaviours

The process of selecting correct gaze behaviours for a speaker during an conversation is an essential part in the evaluation of the conversation between the speaker and the listener. During this interaction the embodied agent has the possibility to gaze towards or gaze away from its interlocutor. Besides selecting the correct gaze behaviours, the gaze frequency and the gaze duration have to be determined. Two different gaze models will be discussed. Both models have elements that will be used in the formalization of a gaze model for the embodied storyteller.

Gaze Movement Model by Fukuyama et al.

Fukuyama et al. [FO02] created a gaze movement model based on the concept of impression management. Impressions are the images we have of other persons. They are formed through displayed nonverbal behaviours, utterances, appearances and the reputation of the other person. By using impression management people try to influence the perception or impression the another has of that person. As an example Fukuyama et al. describe a salesperson, whose goal it is to make a sale. This goal is never told verbally, but by using social control (influence the behaviours and attitudes of others) and impression management the salesman tries to create trust and keep the attention of the client and give the client the impression he would never deceive the client. If an embodied agent is aware of the impressions it is currently bringing across to the user, than the agent can influence the expectations of the user. The resulting impression the user has of the embodied agent controls the resulting behaviours and attitudes of the user [FO02]. To control the impressions an embodied agent brings across to the interlocutor, Fukuyama et al. describe a gaze movement model. The model outputs gaze points that are derived from three different gaze parameters [FO02]. Picked from a large number of psychological studies, each parameter is a statistical variable that is expected to have effect on the impression that is formed by the user:

- Amount of gaze (R): Percentage of the total interaction in which the agent gazes at the user
- Mean duration of gaze (L): Average length of time in which its gazes at the user
- Gaze points while averted (P): A region that describes gaze points other than directly in the users eye

To evaluate the gaze movement model, Fukuyama et al. created an experiment. In this experiment participants described the different impressions they felt when viewing certain gaze patterns. The values used for these gaze patterns are for the three parameters. These values included the standard values derived from existing literature and some extreme values. Table 1 shows the different values used in the experiment. Where R_0, L_0 and P_0 is the standard value for the parameters found in the existing literature and the -, + and ++ are respectively the lower and higher more extreme values. For the gaze points while averted P_0 is the random gaze set and P_H is to gaze up, P_L is to gaze down and P_R is a gaze towards the right (figure 12).

Gaze Parameters	Parameter Values
Amount of Gaze: *R [−]*	$R_- = 0.25$, $R_0 = 0.5$, $R_+ = 0.75$, $R_{++} = 1.0$
Mean Duration of Gaze: *L [ms]*	$L_- = 500$, $L_0 = 1000$, $L_+ = 2000$
Gaze Points while Averted: $P \subset \{(x, y) \mid -\infty \leq x, y \leq \infty\}$	$P_0 = \{(x, y) \mid -1.2 \leq x, y \leq 1.2\}$, $P_H = \{(x, y) \mid -1.0 \leq x \leq 1.0, 0.7 \leq y \leq 1.1\}$, $P_L = \{(x, y) \mid -1.0 \leq x \leq 1.0, -0.9 \leq y \leq -1.3\}$, $P_R = \{(x, y) \mid -2.0 \leq x \leq -1.2, -0.4 \leq y \leq 0.4\}$ *A value determined by a normalization function, based on the user's face region.*

Table 1: Values of Gaze Parameters [F005, p45]

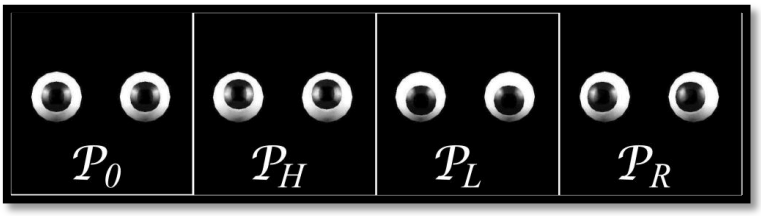

Figure 11: The eyes-only agent [FO02, p46]

In each session the subject first viewed a recorded conversation of an eyes-only agent (figure 11) with the standard gaze parameter conditions, followed by a recorded conversation that used the lower respectively higher values for each of the three parameters. The impressions measured were based on social psychological studies. These studies used ratings that have to do with friendliness (friendly vs. unfriendly, warm vs. cold, sociable vs. unsociable, etc) and dominance (strong vs. weak, successful vs. unsuccessful, careful vs. careless, etc). Figure 12 shows the valuation for the different parameters and their different values.

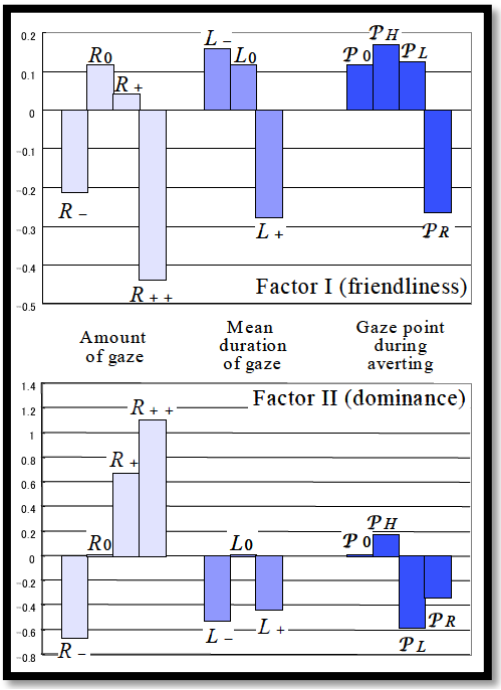

Figure 12: Scores of impression factors [FO02, p47]

Fukuyama et al. formed several of hypotheses regarding the different gaze parameters. Of five hypotheses formed only two were supported by the results. The first hypothesis supported by the results, stated that shorter gaze durations (L_-) yield lower ratings for the agent in impressions measured related to "strong". The agent with a shorter gaze duration at the participant, rated as weaker compared to an agent with average gaze durations (L_0). The other hypothesis that was supported, stated that gaze points for which the agent is gazing downwards (P_L), yields lower ratings in impressions measures related to "strong"'. The agent gazing downwards is rated as weaker, compared to an agent with the random set of averted gaze points (P_0).

Human-like Gazing Model by Mutlu et al.

Mutlu et al. [MF06] describe a gaze model that represents human-like gazing behaviour during storytelling. They use an existing algorithm proposed by Cassell et al. [TJ97], that serves as a starting point for their own model. The algorithm uses the themes and rhemes of an utterance of an English sentence to determine the gaze direction of the speaker. From empirical data Cassell et al. found that the speaker gazes away from the listener at the beginning of a theme with an seventy percent probability and the speaker gazes towards the listener with a seventy-three percent probability at the beginning of a rheme. Cassell et al. suggested the following algorithm to simulate the gaze of a speaker.

```
for each proposition do
    if proposition is theme then
        if beginning of turn or distribution(0.70) then
            attach a look-away from the listener
        end if
    else if proposition is rheme then
        if end of turn or distribution(0.73) then
            attach a look-towards the listener
        end if
    end if
end for
```

Algorithm for gaze behaviour assignment [TJ97, p08]

One remark that can be made about proposed algorithm by Cassell et al. is the absence of selecting gaze behaviours when the *beginning of turn or distribution(0.70)* or *the end of turn or distribution(0.73)* returns false. What kind of gaze behaviour does the speaker select during these situations? Does the agent continue with the current selected gaze behaviour? Or does the agent select gaze behaviour that is opposite of the *look-towards the listener* or *look-away from the listener* that would have been selected?

Mutlu et al. [MF06] extended the algorithm of Cassell et al. [TJ97] with data collected from the recordings of a professional storyteller. These annotations resulted in specific gaze locations and their gaze frequencies. Mutlu et al. discovered four distinct gazing locations. The first and the second location are the two listeners who formed the audience. The third location was a fixed spot on the table in front of the storyteller. The fourth gazing points were a set of random locations in the room. The results of the different gazing locations can be found in table 2.

	Listener 1	**Listener 2**	**Fixed Spot**	**Random spot**
Frequency(%)	13	11	38	38
Length (%)	38	27	30	5
Min (ms)	477	484	242	360
Max (ms)	15,324	5,914	13,674	4,383
Mean (ms)	2,400	2,262	2,640	1,072
Approx. StDev. (ms)	500	500	500	250

Table 2: Length and distribution of each gaze point [MF06, p520]

Mutlu et al. [MF06] used the "Looking at" and "Looking away" state derived from Cassell et al. [TJ97] for the description of their own model with the four distinct gaze locations. They described "Looking at" as keeping the agents gaze towards the listener once it was fixated there. "Looking away" is described as gazing towards the other listener, a random spot or a fixed spot. For the situation that the agent was not currently "Looking at" one of the listeners, "Looking at" meant gazing towards one of the listeners and "Looking away" meant gazing at one of the four gaze targets with a predetermined probability. These probabilities were derived from the frequencies in which the agent gazed at one of the four target locations (table 2).

Figure 13: Asimo telling a story to two listeners [MF06]

The gaze model for human-like gazing of an storyteller was implemented in Honda's humanoid robot, called ASIMO (figure 13). The gaze model used a hand-coded script with markings for the themes, rhemes and pauses of an utterance. Besides the gaze behaviours, ASIMO had the possibility to perform ten simple hand gestures and recited a Japanese fairy-tale. The pseudo-code of the human-like gazing model can be found on the next page.

```
for each part of the utterance (theme/rheme/pause) do
   while the duration of the part do
       if current part is pause then
          if distribution(probability_randomSpot)) then
             gaze at random spot with length(randomSpot)
          else
             gaze at fixed spot with length(fixedSpot)
          end if
       else if current part is theme then
          if distribution(0.70) then
             if distribution(probability_randomSpot) then
                gaze at random spot with length(randomSpot)
             else
                gaze at fixed spot with length(fixedSpot)
             end if
          else
             if distribution(probability_listener1)) then
                gaze at listener1 with length(listener1)
             else
                gaze at listener2 with length(listener2)
             end if
          end if
       else if current part is rheme then
          if distribution(0.73) then
             if distribution(probability_listener1)) then
                gaze at listener1 with length(listener1)
             else
                gaze at listener2 with length(listener2)
             end if
          else
             if distribution(probability_randomSpot) then
                gaze at randomSpot with length(randomSpot)
             else
                gaze at fixedSpot with length(fixedSpot)
             end if
          end if
       end if
   end while
end for
```

The version of the gaze model found in the paper of Mutlu et al. only returned gaze behaviours at random spots with a gaze duration determined by one of the available gaze locations (listener1, listener2, fixedSpot, randomSpot). It can be assumed that Mutlu et al. meant that when a duration of the certain gaze location is chosen, also the gaze location itself would be selected as the next gaze point. The above model has been corrected because of this.

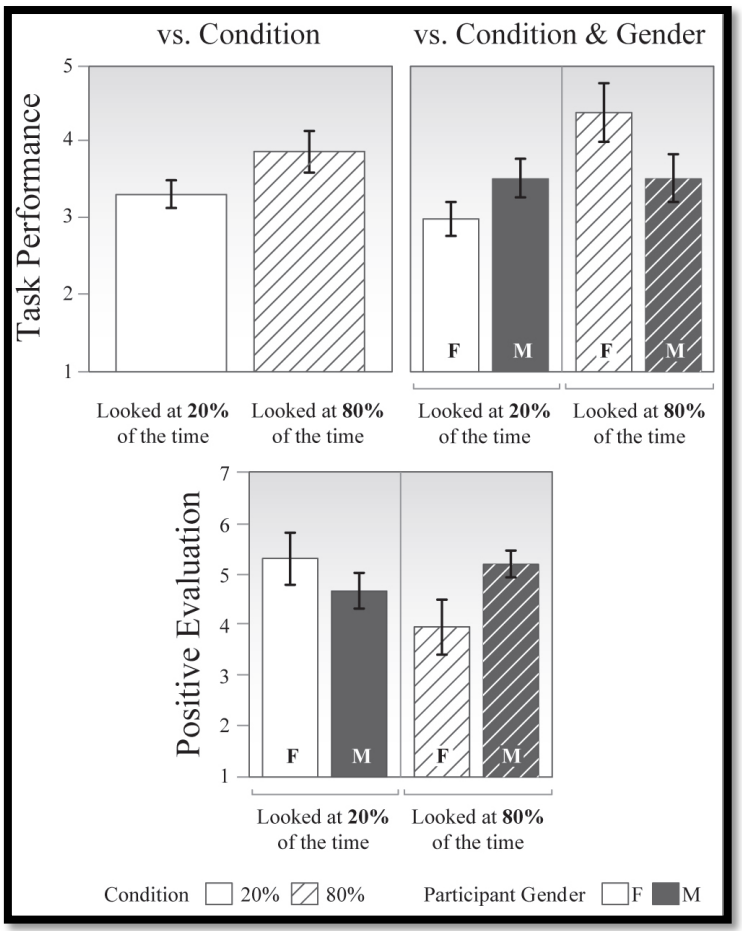

Figure 14 Top: Main effect of condition and interaction between condition and participant gender on task performance. Bottom: Interaction between condition and participant gender on positive evaluation of the robot [MF06, p522]

As mentioned before the human-like gaze model was incorporated into Honda's humanoid robot ASIMO, which told a Japanese story towards an audience. Mutlu et al. tried to evaluate the following two hypotheses:

- Participants who are looked at more will perform better in the recall task than participants who are looked at less
- Participants who are looked at more will evaluate ASIMO more positively than participants who are looked at less.

The gaze of ASIMO was manipulated during an experiment to evaluate both hypotheses. In this experiment ASIMO gazed at one of the participants only twenty percent of the time and the other participant was gazed at eighty percent of the time. The experiment started off with ASIMO presenting himself and then performed a storytelling task. After listening to ASIMO the participants had to listen to another story on tape, which was called the distracter task. Beforehand they were told that they had to answer questions about one of the presented stories.

Pre,- and post-questionnaires were used to assess the model. The results from these questionnaires can be found in figure 14. The results supported the first hypothesis of Multu et al., that participants who are looked at more will perform better. The second hypothesis was not completely supported, because it was seen that gender played a role in the evaluations. Women tend to evaluate ASIMO more positively if they were looked at less.

3.3 Discussion

If people are interacting with each other they often use nonverbal behaviours to augment the experience. The behaviours provide a rich source of information and each serve a specific function during the (social) interaction. As mentioned in the introduction gaze behaviours were chosen as the main nonverbal behaviour of the Embodied Storyteller. For example head movements require too much context to make it possible to create a behavioural model without speech or text analysis. So the focus of this thesis will be on the head movements that accompany a gaze behaviour.

The gaze models discussed described the different elements involved in forming correct gaze behaviours. Fukuyama et al. [FO02] showed that variances in gaze frequency and gaze duration influence the impression a user has of an agent. Mutle et al. [MF06] demonstrated that people who are looked at more often will better recall a task. The Embodied Storyteller will use different elements of each model for its own gaze model. First we need to determine the correct gaze frequencies, gaze durations and the available gaze points for Embodied Storyteller during the telling of a story. To determine these elements data will be collected by annotation an actual storyteller. The annotation procedure and the results will be discussed in the upcoming chapter.

4 DATA COLLECTION

Two viewing perspectives are investigated, namely the newsreader and storyteller viewing perspective. Specific gaze models for both viewing perspectives will be created to determine the suitable gaze behaviours at certain times in the story. Several video sessions of an actual storyteller have been annotated to observe the variety in different gaze behaviours of an actual storyteller. The annotations not only resulted in the description of different gaze behaviours at specific gaze points, but also the specific gaze frequencies and gaze durations have been derived.

For the storyteller perspective, with an camera placement besides an audience, two annotation iterations were performed. In the first annotation iteration ten video fragments were annotated of an actual storyteller. This first iteration annotated the gaze behaviours in seconds. It became apparent that after reviewing the results of the first iteration, that perhaps annotating the behaviours in milliseconds would be better. Therefore a second annotation iteration was introduced and the professional annotation software ELAN (Eudico Linguistic Annotator, http://www.lat-mpi.eu/tools/elan/) was used for this iteration. For the newsreader perspectives, in which the storyteller talks directly towards the camera, the gaze behaviours are a subset of the gaze behaviours found in the storyteller. The only difference being the absence of an audience. For the gaze frequencies and the gaze durations for the storyteller in the newsreader perspective, two news broadcast session were annotated using ELAN. In the upcoming sections the results of the annotations for both perspectives will be provided and the findings and implications of the results will be discussed.

4.1 First Annotation Iteration of an actual Storyteller

What are the minimum requirements of the video material that will be annotated? The first requirement is a clear and prolonged view of the storyteller. This ensures that the different gaze behaviours can be annotated from start to end. The second requirement is a minimum amount of video material of the actual storyteller to see if any common gaze behaviours emerge. The final requirement is the existence of an audience. The audience introduces a gaze behaviour not found in a regular face-to-face interaction. This added gaze behaviour can have a positive effect on to the overall viewing experience and should therefore be tested.

Used Material

Most of the requirements are met by the Dutch TV Series "Elly en de wiebelwagen". The series started in 2006 and is still running. It is being broadcasted by the "EO (Evangelische Omroep)" on a public broadcasting channel. "Elly en de wiebelwagen" tell stories from a evangelical perspective. Other alternatives, like for example Sesame Street, do not have public available episodes because of copy-right infringements. The gaze behaviours displayed by the actual storyteller, is not the general gaze behaviour of *all* storytellers. The annotation of the gaze behaviours, the frequency and the duration of the gaze serve as a starting point for the possible gaze behaviours the embodied storyteller displays. The objective is not to determine general gaze behaviours of all storytellers, the objective is to create interesting viewing experiences from the different perspectives and the gaze behaviours that fit within these viewing perspectives. Thus the analysis of the gaze behaviour of single storyteller should suffice.

Figure 15: On the left shows Elly in her wobbly wagon, on the right Dop and Kurk sitting behind an audience.

Procedure

The storyteller from "Elly en the Wiebelwagen" is Elly (figure 15), she travels around in her wobbly wagon and tells stories to children. The children form a small audience, together with Dop and Kurk (figure 15). Elly tells her story from a large book, which she receives from Dop and Kurk. Dop and Kurk are two male characters that travel around with Elly and sit behind the children. Some parts of the story are visualized in a drawing (figure 16), this happens during the telling of the story by Elly. This means that there are moments in the video in which the storyteller is not visible. In these moments the story is either visualized in a drawing or the audience is visible. The assumption is made that the gaze behaviours of the Elly continue and that these gaze behaviours are no different than the visible gaze behaviours.

Figure 16: Visualization of the story in a drawing

The first ten episodes of "Elly en de wiebelwagen" were chosen to be annotated. The length of each episode can be found table 3. In each of the sessions the gaze behaviours are annotated from the start to the end. Table 2 shows a fragment of the annotation of one episode. Also the moment in which the audience is visible and the visualization of the story in a drawing are written down from beginning to end.

Episode	**Duration (hh:mm:ss)**
2008-01-09	0:02:58
2008-01-16	0:02:38
2008-01-30	0:02:54
2008-02-06	0:02:26
2008-02-13	0:03:18
2008-02-20	0:03:12
2008-02-27	0:02:43
2008-03-05	0:02:40
2008-03-19	0:03:11
2008-04-09	0:02:31
Average length	**0:02:51**

Table 3: The annotated episodes with their individual length

BEGIN	**END**	**ANNOTATION**
0:29	0:30	Gazing towards the left-side of the book
0:30	0:31	Gazing towards the audience, moving posture slightly back, slight head shake
0:31	0:32	Gazing towards the left-side of the book
0:32	0:33	Gazing towards the audience
0:33	0:36	Shot of the audience
0:36	0:37	Gazing towards the audience, moving gaze from right to left, slight nod
0:37	0:39	Gazing towards the left-side of the book
0:39	0:45	Visualization of the story in a drawing
0:45	0:46	Gazing towards the audience, head slightly tilted sideways, moving posture forwards
0:46	0:46	Gazing towards the left-side of the book
0:46	0:52	Shot of two actors sitting behind the audience

Table 4: Short section of an annotation of "Elly en de Wiebelwagen (02-20-2008)"

Available Gaze Behaviours

The annotations of the first ten episodes of "Elly en de Wiebelwagen" provided five different gaze behaviours. They reoccur that often, that they were selected as the main gaze behaviours of the embodied storyteller. Of these five gaze behaviours, the frequency of the gaze and the durations of the gaze was written down. This information will be used in the construction of a gaze model for the embodied storyteller. The specific gaze behaviours are:

- Gaze at audience
- Gaze at the left page of the book
- Gaze at the right page of the book
- Gaze at the camera
- Gaze story related

Story related gazing is for example gazing at the ceiling when she is pointing at the sun is or when she holds an object in her hand and she gazes at that object.

Results

The average length of each episode is around two and a half minutes (table 2). In this time the five main gazing behaviours of the actual storyteller are written down. From the ten episodes the frequency of the gaze, the average duration of a gaze as well as the percentage of the gaze behaviour was written down.

Table 4 shows the frequency of the different gaze behaviours. The results show that on average the actual storyteller gazes most often at the audience or at one of the pages of the book. Table 5 shows the average duration of the gaze behaviours. This averages around one second per gaze behaviour. The average percentage of gaze behaviours during an episode can be found in table 6. Again it shows that most of the gazes occur at either the audience, the left page of the book or at the right page of the book.

Gazing behaviour	**Gaze frequency (per minute)**
Gaze at the audience	21
Gaze at the right-page of the book	13
Gaze at the left-page of the book	11
Story related gaze	2
Gaze at the camera	5

Table 5: Gaze frequency

Gazing behaviour	**Mean duration in seconds**
Gaze at the audience	1.00
Gaze at the right-page of the book	1.00
Gaze at the left-page of the book	1.00
Story related gaze	1.00
Gaze at the camera	1.00

Table 6: Mean duration of the gaze behaviour

Gazing behaviour	**Amount of gaze**
Gaze at the audience	44%
Gaze at the right-page of the book	21%
Gaze at the left-page of the book	19%
Story related gaze	5%
Gaze at the camera	11%

Table 7: Amount of gaze

4.2 Second Annotation Iteration of an actual Storyteller

Annotating the different gaze behaviours in seconds was not the most accurate method of writing down the gaze behaviours. To verify the found results three episodes of "Elly en de Wiebelwagen", that were also annotated in the first iteration, were annotated in the professional annotation software ELAN (http://www.lat-mpi.eu/tools/elan/). The results from ELAN will be discussed in the upcoming sections.

Used Material

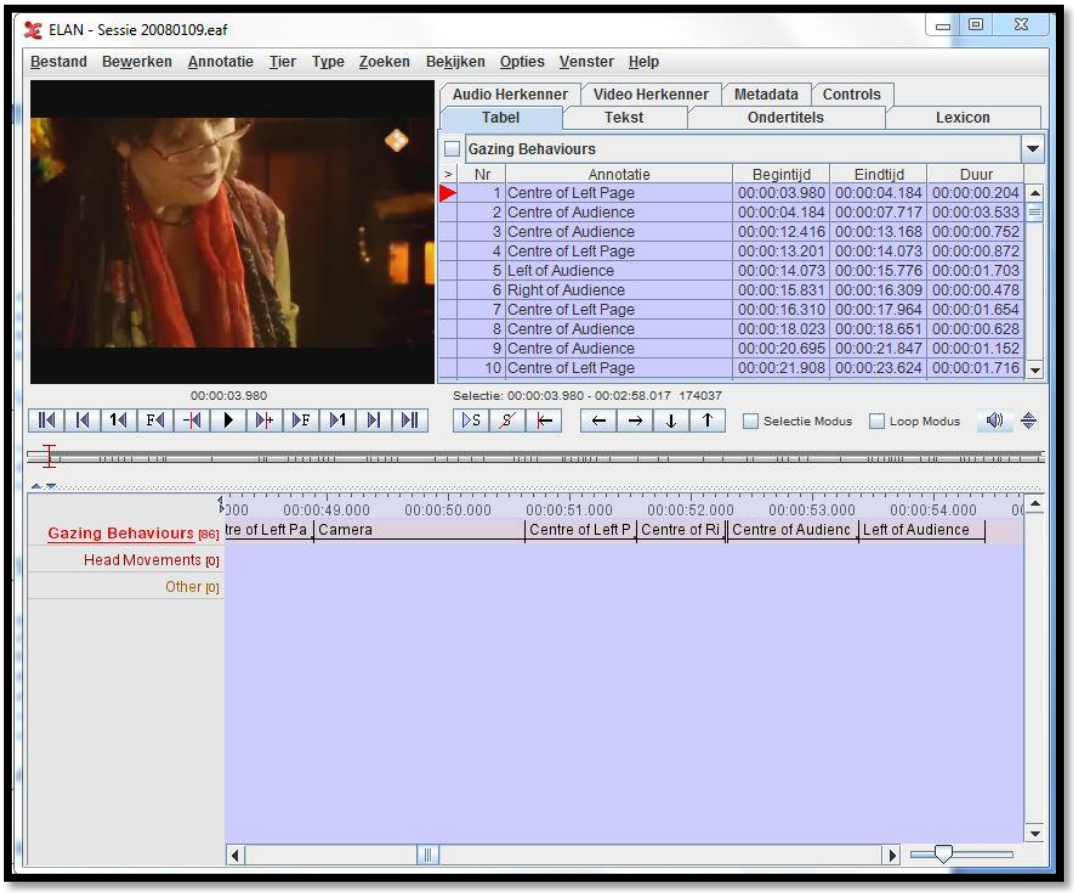

Figure 17: ELAN

ELAN is a professional tool for the creation of complex annotations (http://www.lat-mpi.eu/tools/elan/). It enables the user to annotate a video fragments in millisecond and use specific layers to separate the annotations per category. ELAN also provides statistical information after the annotations have been completed. Figure 17 provides an overview of the used software.

Procedure

Three layers (Tiers) have been created representing gaze behaviours, head movements and other parts. In the top left of ELAN the current episode of "Elly en de Wiebelwagen" that is annotated can be viewed. The top right of ELAN shows the annotated gaze behaviours with their start and end. Below shows the section were annotation labels are added.

As can be seen in figure 17 three layers (Tiers) were created. During the course of the annotations it was decided to put the focus on the *Gaze Behaviours* and omit *Head Movements* from the annotations in ELAN. Head movements require too much contextual information and would be difficult to successfully implement in the embodied storyteller with the amount of time available. The layer *Other* refers to situations where either the story is visualized in a drawing or the audience is visible. Both situations can be derived from the first annotation iteration and do not need to be calculated in milliseconds.

Annotations proceeds in ELAN by processing the time line of an episode in the bottom half of ELAN. Every instance of a gaze behaviour has a start and an end and is labelled with its type of gaze behaviour. Gaze behaviours are predetermined and are derived from an available vocabulary linked to a layer.

Available Gaze Behaviours

The gaze behaviours are split-up into more specific gaze behaviours. The gaze at the audience was split up into three different gaze behaviours, namely gaze at the left, centre or right-side of the audience. The designation of these gaze points can be seen as a comment on the fact that within the audience the storyteller gazes at specific individuals. Subsequently the book was divided into six different gaze behaviours (starting at the top, middle or bottom of either the left of right page). However during the annotations in ELAN it became apparent that it was too hard to distinguish between the transitions from top to centre and from centre to the bottom of a page. The distinction between these points was therefore omitted and only the gaze at either the centre-left or centre-right page remained.

Results

Table 8, 9 and 10 show the average results of the three annotated video fragments. Table 8 again shows a trend in which most of the gaze behaviours occur at either the audience, the left page of the book or the right page of the book, unlike the gaze at the camera and the story related gaze that occur less often. Table 9 shows that the average duration of the gaze behaviour is still around one second with the story related gaze being the highest with 1.520 seconds and the gaze at the right-side of the audience the lowest with 0.895 seconds. Finally there is the average percentage of the gaze behaviour per minute, as with the average amount of gaze behaviours, with the gaze at the audience and the gaze at the left or right page at the book being the main gaze behaviours.

Gazing behaviour	**Gaze frequency (per minute)**
Gaze at the audience	32
Gaze at the right-page of the book	15
Gaze at the left-page of the book	12
Story related gaze	1
Gaze at the camera	3

Table 8: Gaze frequency

Gazing behaviour	**Mean duration (seconds, milliseconds)**
Gaze at the audience	0.970 (0.7)
Gaze at the right-page of the book	1.110 (0.5)
Gaze at the left-page of the book	1.074 (1.0)
Story related gaze	1.520 (0.8)
Gaze at the camera	1.375 (0.6)

Table 9: Mean duration of the gaze behaviour, standard deviation in parenthesis

Gazing behaviour	**Amount of gaze**
Gaze at the audience	45%
Gaze at the right-page of the book	18%
Gaze at the left-page of the book	22%
Story related gaze	9%
Gaze at the camera	6%

Table 10: Amount of gaze

4.3 Annotation of a News Broadcast

Different viewing perspectives can create different viewing expectations. Part of these expectations have to do with the gaze duration and gaze frequency at one of the available gaze points. As the name implies, the newsreader perspective is based on a camera perspective found in a news broadcast. The newsreader talks and gazes directly towards the camera and sometimes gazes at a piece of paper in front of him (see figure 18). Too see how the duration and gaze frequency of the newsreader compares to that of the actual storyteller, two news broadcasts were annotated.

Used Material

The news that is broadcasted by the NOS (Nederlandse Omroep Stichting) was used as the video material to be annotated. The NOS is a news broadcaster on the Dutch public broadcasting channels. One of its main tasks is to provide current news and sports coverage. The news items chosen last around fifteen minutes. In this time the newsreader gazes at the camera and at a piece of paper he is holding in his hand multiple times (figure 18).

As with the storyteller, there are moments in the news broadcast were the newsreader is not visible. In this time the story is visualized with background material. To see if any common gaze frequencies or gaze durations emerge, the same news anchor was used in both video fragments.

Figure 18: News broadcast aired 2011-11-14 by the NOS

Procedure

ELAN was used as the annotation tool to determine the gaze frequency and gaze duration of the newsreader. The annotation layers of the previous annotations of the actual storyteller, were also used for the annotations of the news broadcasts.

Available Gaze Behaviours

Three main gaze behaviours are displayed by the newsreader. The first gaze behaviour is the gaze at the piece of paper, which he holds in his hands or lies on the table. The second gaze behaviour is at the camera and is positioned directly in front of the newsreader (see figure 18). Finally there is gaze the behaviour called news item gazing. This gaze behaviour occurs when the newsreader gazes at a video monitor (positioned at the lower left part of the screen) when a video fragment with the current news item ends and the newsreader becomes visible again.

Results

The results of gaze behaviours displayed by the newsreader can be found in table 11, 12 and 13. The tables show averages of the different gaze frequencies, gaze durations and the average percentage of the different gaze behaviours per minute. The results demonstrate that the newsreader gazes six times per minute at the camera lasting on average around ten seconds. The gaze at the piece of paper and gaze at the monitor displaying the current news item occurs two and three times a minute. The gaze at the piece of paper lasts around a second and the gaze at the monitor 0.4 seconds.

Gazing behaviour	**Gaze frequency**
Gaze at the camera	6.41
Gaze at a piece of paper	2.48
Gaze at the news item	2.76

Table 11: Gaze frequency

Gazing behaviour	**Mean duration (seconds, milliseconds)**
Gaze at the camera	10.2 (5.5)
Gaze at a piece of paper	1.0 (0.1)
Gaze at the news item	0.4 (0.5)

Table 12: Mean duration of the gaze behaviour, standard deviation in parenthesis

Gazing behaviour	**Amount of gaze**
Gaze at the camera	94%
Gaze at a piece of paper	4%
Gaze at the news item	2%

Table 13: Amount of gaze

4.4 Discussion

The first annotation iteration and the second annotations in ELAN show some differences in the results for both the frequency of the gaze and the duration of the gaze. Regarding the frequency of the gaze the most common gaze behaviours remain the same. These gaze behaviours are either at the left-page or right-page of the book or at the gaze at the audience. The duration of the different gaze behaviours demonstrate the biggest differences between the two annotation iterations, especially with the gaze at the camera and story-related gazes. Instead of every gaze behaviour lasting around one second, the gaze at the camera should have an average gaze duration of 1.4 seconds and the story related gazes should have an average gaze duration of 1.5 seconds.

Subsequently the annotations of the two newsreader episodes demonstrate that the gaze at the camera, for both its frequency and duration, is more frequently and much longer than that of the actual storyteller from " Elly en de wiebelwagen". The video fragments are difficult to compare. The results of the newsreader annotations should be used as indication of gaze values used when constructing the gaze model. Perhaps adjustments in the gaze frequency and gaze duration for the embodied storyteller from a newsreader perspective, better meet the expectations of the user in the performed behaviours of the embodied storyteller. How do the results found in the annotations compare to the data used in existing gaze models discussed in the literature section?

Let's start with the gaze movement model by Fukuyama et al. [FO02]. The gaze movement model creates gaze behaviour for an (eyes-only) agent that talks directly towards the user. Fukuyama et al. gathered data from existing literature to determine the different gaze frequencies and gaze durations. They made a distinction between the gaze at the user and the gaze points while averted. The agent gazed half of the time at the user and the other half of the time at gaze points while averted. The gaze duration lasted on average around one second for every available gaze point. When creating a gaze model from a newsreader perspective, a choice has to be made to use either the results found in the face-to-face interaction in the gaze movement model, or the results found in the annotations from the news broadcast sessions. The choice of used gaze durations and gaze frequencies for the newsreader perspective will be answered in the upcoming section, when the gaze model for the embodied storyteller is created.

Next there is the human-like gazing model by Mutlu et al. [MF06]. The model used data derived from the annotations of a real-world storyteller. Four gaze locations were described, namely two listeners, a fixed spot on the table and a set of random spots. The frequencies of the gaze and the gaze duration are quite different compared to the results found in the annotations of the storyteller from "Elly en die wiebelwagen". If these differences are due to the characteristics of the storyteller, the situation or some other factor is hard to tell. It is only reasonable to use the gaze durations and frequencies for the annotations of "Elly en de wiebelwagen", because the situation from "Elly en wiebelwagen" will be recreated in a virtual environment.

5 GAZE MODEL

A gaze model will be created to simulate the gaze behaviours of an actual storyteller. The gaze behaviours are derived from the annotations of "Elly en de wiebelwagen" and the news broadcasts. They are combined with the information gathered from the existing literature. Let's start with the parameters provided by Fukuyama et al. [FO02]. They will serve as a starting point for the gaze model of the embodied storyteller. The parameters are the amount of gaze, the duration of the gaze and the gaze points while averted, where the final parameter will be reformulated as a parameter that provides specific gaze points at certain times. The model of Mutlu et al. [MF06] based on the algorithm provided by Cassell et al. [TJ97] returns specific gaze points that depend on the distributions for the themes and rhemes of an utterance. The distributions determine the look-away and look-at state of the speaker towards the listener. The gaze model of the embodied storyteller will use the gaze distributions derived from the annotations to determine the current "look-at" state for the different gaze points of the embodied storyteller.

Two gaze models will be created; one for the storyteller perspective, which has five distinct gaze points and one for the newsreader perspective, which has two distinct gaze points. Both gaze models are divided into three parts. The first part of the gaze model will determine the available gaze points at a certain time in the story. The second part of the model selects a gaze point using the gaze distributions and selecting a random gaze point. The final part of the model is to determine the duration of the selected gaze point.

5.1 Storytelling Perspective

The situation of the storytelling perspective has five different elements (figure 19). These five elements are the embodied storyteller and the different gaze points. The positioning of the different elements is a recreation of the storytelling situation as found in "Elly en de wiebelwagen". The camera (4) is placed besides an audience (5) and the storyteller (1) uses a book (2) to tell the story from. Finally there are the story-related gazes (3), these are gazes are at specific objects. These objects are either imaginary (like gazing at the ceiling when talking about the sun) or actual objects (like an object she is holding in her hands). They are predetermined gaze points that depend on the context of the current story and formulated beforehand.

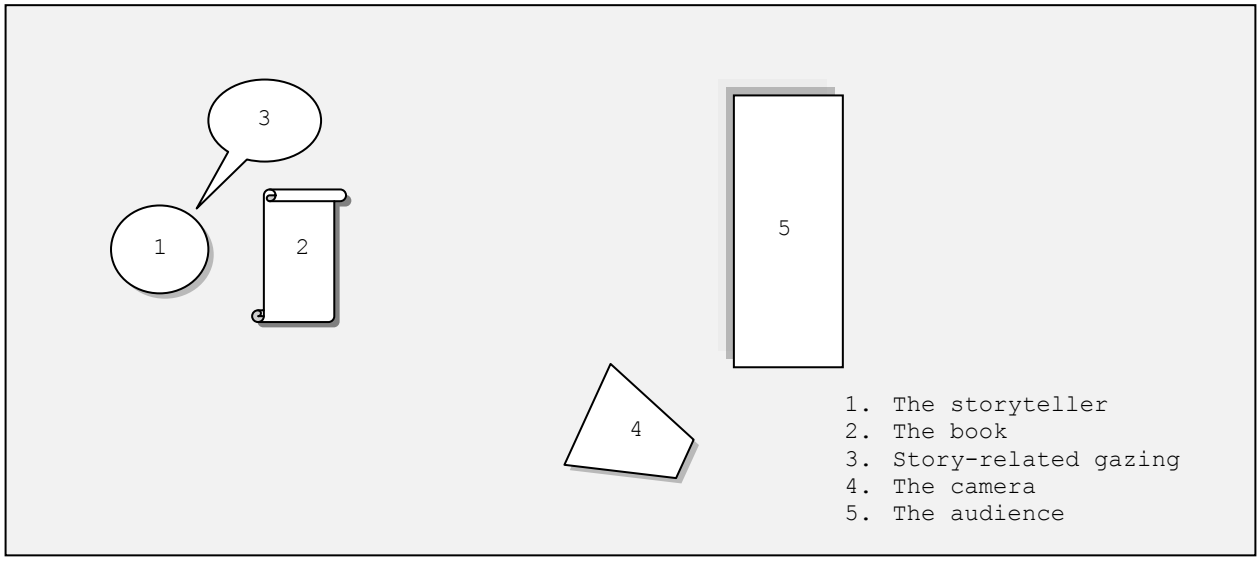

Figure 19: The storyteller perspective

Available Gaze Points

The first part of the gaze model determines the gaze points that are available at a certain time in the story. It is a relatively simple selection process that depends on a number of conditions. These conditions are derived from observations made during the annotations of the storyteller from "Elly en de wiebelwagen":

- Condition 1: When at the beginning of a story, gaze at the left-page of the book. The beginning of "Elly en de wiebelwagen" always started with Elly receiving the book from Dop or Kurk. Subsequently Elly opened the book, gazed at the left page of the book and started telling the story. Because it is too complex to recreate this entire opening sequence, the embodied storyteller will start with a gaze at the left-page of the book.
- Condition 2: When at the end of a story, gaze at the camera. "Elly en de wiebelwagen" always ended with Elly giving the book to Dop and Kurk. It is not feasible to recreate this sequence of events, but a correct ending event for the embodied agent should be formulated. This will be a gaze at the camera.
- Condition 3: The story-related gaze points are determined in advance depending on the context of the story.
- Condition 4: The gaze point to be selected as a next gaze point cannot be the current selected gaze point.
- Condition 5: Determine if the embodied agent is currently reading from the left or right-page of the book. The agent always starts with a gaze at the left-page of the book. After a set time the gaze will transition to the right-page of the book and vice versa.

Algorithm 1 (see next page) shows the gaze model for the embodied storyteller and the algorithm start with a description how available gaze points at a certain time are selected. The above conditions are used to select the available gaze points at a certain time in the story. Time is in seconds and is the accumulation of the duration of the selected gaze behaviour and a transition time. The transition time (with time in seconds) is the duration were the agent moves its head and eyes towards the newly selected gaze point. This transition time is not part of the duration of the gaze, which is the time were the agent actually gazes at the selected gaze point. Subsequently there is the page counter, this is a number that counts how often the left page has been gazed at by the embodied storyteller. After a fixed number of times, derived from the annotated sessions, the gaze at the left-page of the book transitions from the left-page to the right-page and vice-versa.

When the available gaze points have been determined for a certain time in the story, two other methods are performed. The first method, named *nextGazePoint*, selects the next gaze point. The other method, named *selectGazeDuration*, selects the duration of the gaze. Both methods will be described in upcoming sections. Finally the selected next gaze behaviour from the available gazes and its gaze duration are stored, so that it can be used by the embodied storyteller when it is actually performing the gaze behaviours. Finally in the algorithm the gaze points are described with their abbreviations, see table 14 for a description.

The gazing points	**Abbr.**
Gaze at the camera	C
Gaze at the audience	A
Gaze at the left-page of the book	LP
Gaze at the right-page of the book	RP
Story related gaze	S

Table 14: Abbreviations of the gaze behaviours

```
/**
 * Algorithm to determine the available gaze points at a certain time in the story
 **/

currentGazePoint = “”  //start with an empty gaze situation
currentPage = LP  //start reading from the left-page of the book
pageCounter = 0 //time passed since last page transition
time = 0 //start at the beginning of the story (time is in seconds)
gazeScriptForStory = {}
WHILE time < story.length DO

      /** Determine the available gaze points **/
      availableGazePoints = {}
      IF time EQUALS 0 THEN
            availableGazePoints = {LP}
      ELSE IF time EQUALS story.length THEN
            availableGazePoints = {C}
      ELSE IF NOT(currentGazePoint.equals(S)) AND story requires S at time T THEN
            availableGazePoints = {S}
      ELSE
            IF NOT(currentGazePoint.equals(C)) THEN
                  availableGazePoints.addElement(C)
            END IF

            IF NOT(currentGazePoint.equals(A)) THEN
                  availableGazePoints.addElement(A)
            END IF

            IF NOT(currentGazePoint.equals(LP)) AND currentPage.equals(LP) THEN
                  IF pageCounter < time to read a page THEN
                        availableGazePoints.addElement(LP)
                  ELSE
                        availableGazePoints.addElement(RP)
                        currentPage = RP
                        pageCounter = 0
                  END IF
            ELSE IF NOT(currentGazePoint.equals(RP)) AND currentPage.equals(RP) THEN
                  IF pageCounter < time to read a page THEN
                        availableGazePoints.addElement(RP)
                  ELSE
                        availableGazePoints.addElement(LP)
                        currentPage = LP
                        pageCounter = 0
                  END IF
            END IF
      END IF

      /** Select a gaze point from the available gaze points **/
      nextGazePoint = selectGazePoint(availableGazePoints)

      /** Determine the correct gaze durations **/
      durationNextGaze = selectGazeDuration(nextGazePoint)

      /** Increment the time with the duration of the gaze and the transition time **/
      time = time + durationNextGaze + gazeTransitionTime

      /** Add the time of the duration of the next gaze and the transition time **/
      pageCounter = pageCounter + durationNextGaze + gazeTransitionTime

      /** Store the selected gaze points and gaze lengths **/
      gazeScriptForStory.add([nextGazePoint, durationNextGaze])

END WHILE
```

Algorithm 1: Gaze model for the embodied storyteller from a storytelling perspective

Selecting the Next Gaze Point

The next step in the algorithm is to select one gaze point from the available gaze points. Table 15 shows the amount of gaze of the storyteller from "Elly en de Wiebelwage" at the different gaze points. The amount of gaze during an entire story will be used to determine the chance that a certain gaze point is selected from the available gaze points. An equal distribution of gaze elements is created to randomly select one of the available gaze points. The distribution of the gaze elements, depends on the average amount of gaze at the different gaze points. From this gaze distribution a random element is selected. The selected element is the gaze point that will be used as the next gaze point for the embodied storyteller.

The gazing points	**Amount of gaze**
Gaze at the audience	45%
Gaze at the right-page of the book	18%
Gaze at the left-page of the book	22%
Story related gaze	9%
Gaze at the camera	6%

Table 15: Amount of gaze for the storyteller perspective

Algorithm 2 describes the selection of a random gaze point from a distribution of gaze elements. The algorithm starts with selecting a random number (rn) from 100 elements. *beginAt* represents the starting element for the different intervals of the different gaze elements. For example the interval starting at zero and ending at five, represents the six percent of elements that represent the gaze at the camera (table 15). Usually not all gaze points are available as a next gaze point and the amount of gaze at a certain gaze points need to be normalized.

The normalization of the amount of gaze will be described in an example. Let's say that the current gaze point is a story-related gaze and the available gaze points that can be selected as the next gaze point are the gaze at the camera, the audience and the left-page of the book. As mentioned before the distribution of elements determines the chance that a certain gaze point can be selected. To select a gaze element, a distribution of one-hundred elements is created. Of these one-hundred elements ((6)/(6+45+22)*100) = 8 percent is the gaze at the camera, ((45)/(6+45+22)*100) = 62 percent is the gaze at the audience and finally ((22)/(6+45+22)*100) = 30 percent is the gaze at the left-page of the book. After the normalization of the amount of gaze per gaze point a random element is chosen from the distribution of elements. The selected element is one of the available gaze points. The gaze point will be the gaze point for time T in the story.

```
/**
 * Algorithm to determine the gaze point that will be selected next
 **/

rn = random(0,99) //select a random element from 100 elements
beginAt = 0;
FOR EACH gazePoint : availableGazePoints DO
    IF gazePoint.equals(C) AND beginAt <= rn < (beginAt + norm(amountOfGaze(C)))THEN
       nextGazePoint = C
       beginAt = beginAt + amountOfGaze(C)
    ELSE IF gazePoint.equals(S) AND beginAt <= rn < (beginAt + norm(amountOfGaze(S)))THEN
       nextGazePoint = S
       beginAt = beginAt + amountOfGaze(S)
    ELSE IF gazePoint.equals(A) AND beginAt <= rn < (beginAt + norm(amountOfGaze(A)))THEN
       nextGazePoint = A
       beginAt = beginAt + amountOfGaze(A)
    ELSE IF gazePoint.equals(LP) AND beginAt <= rn < (beginAt + norm(amountOfGaze(LP)))THEN
       nextGazePoint = LP
       beginAt = beginAt + amountOfGaze(LP)
    ELSE IF gazePoint.equals(RP) AND beginAt <= rn < (beginAt + norm(amountOfGaze(RP)))THEN
       nextGazePoint = RP
       beginAt = beginAt + amountOfGaze(RP)
    END IF
END FOR EACH
```

Algorithm 2: Select a gaze point from the available gaze points for storyteller perspective

Determine Gaze Length

The final step in the algorithm is to determine the duration of the gaze. The duration is a number that lies between the mean of the gaze duration and its standard deviation. The variance in the gaze duration enables us to create a more varied viewing experience, compared to gaze durations that always have the same length. Table 16 shows the results from the annotations of the gaze durations of the storyteller from "Elly en de wiebelwagen". The formalization of the selection of the specific gaze duration can be found in algorithm 3.

Gazing behaviour	**Mean duration (μ)**	**Standard deviation (σ)**
Gaze at the audience	0.970	0.7
Gaze at the right-page of the book	1.110	0.5
Gaze at the left-page of the book	1.074	1.0
Story related gaze	1.520	0.8
Gaze at the camera	1.375	0.6

Table 16: Gaze duration derived from the storyteller annotations

```
/**
 * Interval of values that can be the duration of the gaze
 * Where μ and σ are the values for the selected gaze point found in table 16
 **/

Gaze duration(Gaze Point): [μ - σ, μ + σ] = {x ∈ (μ - σ) ≤ x ≤ (μ + σ)}
```

The set of values for the gaze duration of a specific gaze point for the storyteller perspective

5.2 Newsreader Perspective

The situation of the newsreader perspective consists of four main elements and represents the parts as found in " Elly en de wiebelwagen", with the difference being the position of the camera and the absence of the audience. What remains are the embodied storyteller (1), the book (2), the camera (3) and the (4) story-related gazing. The camera position is derived from the newsreader broadcasts and the results of the annotated sessions of the newsreader serve as an indication for gaze values used by the embodied storyteller.

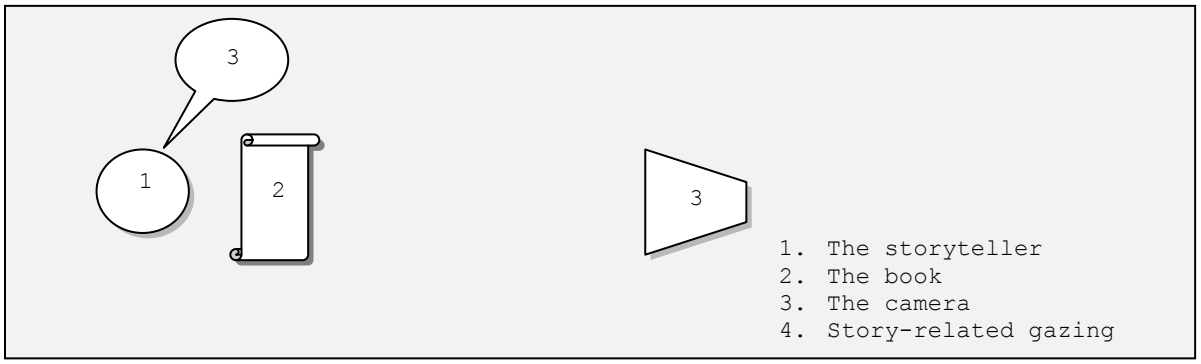

Figure 20: Newsreader perspective

Available Gaze Points

Again there are five conditions that determine the available gaze points at a certain time in the story. These conditions are the same as the storyteller perspective, to summarize:

- When at the beginning of a story, gaze at the left-page of the book.
- When at the end of a story, gaze at the camera.
- The story-related gaze points are determined in advance depending on the context of the story.
- The gaze point to be selected next cannot be the current selected gaze point.
- Determine if the embodied agent is currently reading from the left or right page of the book. The agent always starts with a gaze at the left page of the book, after a set time the gaze will transition to the right page and vice versa.

Algorithm 4 shows the gaze model and the selection process of the available gaze points for the newsreader perspective. It is almost the same as the storyteller perspective, but the gaze at the audience has been removed from the algorithm.

```
/**
 * Algorithm to the determine available gaze points at a certain time in the story
 **/

currentGazePoint = "”  //start with an empty gaze situation
currentPage = LP  //start reading from the left-page of the book
pageCounter = 0 //time passed since last page transition
time = 0 //start at the beginning of the story (time is in seconds)
gazeScriptForStory = {}
WHILE time < story.length DO

      /** Determine the available gaze points **/
      availableGazePoints = {}
      IF time EQUALS 0 THEN
            availableGazePoints = {LP}
      ELSE IF time EQUALS story.length THEN
            availableGazePoints = {C}
      ELSE IF NOT(currentGazePoint.equals(S)) AND story requires S at time T THEN
            availableGazePoints = {S}
      ELSE
            IF NOT(currentGazePoint.equals(C)) THEN
                  availableGazePoints.addElement(C)
            END IF

            IF NOT(currentGazePoint.equals(LP)) AND currentPage.equals(LP) THEN
                  IF pageCounter < time to read a page THEN
                        availableGazePoints.addElement(LP)
                  ELSE
                        availableGazePoints.addElement(RP)
                        currentPage = RP
                        pageCounter = 0
                  END IF
            ELSE IF NOT(currentGazePoint.equals(RP)) AND currentPage.equals(RP) THEN
                  IF pageCounter < time to read a page THEN
                        availableGazePoints.addElement(RP)
                  ELSE
                        availableGazePoints.addElement(LP)
                        currentPage = LP
                        pageCounter = 0
                  END IF
            END IF
      END IF

      /** Select a gaze point from the available gaze points **/
      nextGazePoint = selectGazePoint(availableGazePoints)

      /** Determine the correct gaze durations **/
      durationNextGaze = selectGazeDuration(nextGazePoint)

      /** Increment the time with the duration of the gaze and the transition time **/
      time = time + durationNextGaze + gazeTransitionTime

      /** Add the time of the duration of the next gaze and the transition time **/
      pageCounter = pageCounter + durationNextGaze + gazeTransitionTime

      /** Store the selected gaze points and gaze lengths **/
      gazeScriptForStory.add([nextGazePoint, durationNextGaze])

END WHILE
```

Algorithm 4: Gaze model for the embodied storyteller from a newsreader perspective

Selecting the Next Gaze Point

From the available gaze points a single gaze point is selected. As with the storyteller perspective, this is done by using the amount of gaze from the different gaze points and a distribution of gaze elements. A single element is randomly selected from the distribution and serves as the next gaze point to be gazed at by the embodied storyteller.

From the newsreader annotations it was seen that the amount of gaze at the camera is far higher than the gazes at the camera by the storyteller of "Elly en de wiebelwagen". The most obvious explanation for the difference between the storyteller and newsreader annotations is the absence of an audience. To compensate for this absence of an audience the results from the annotated sessions of the storyteller are adjusted. The amount of gaze at the camera will be the sum of the gaze at the camera (derived from the storyteller annotations) plus the gaze at the audience (derived from the storyteller annotations).

The gazing points	**Amount of gaze**
Gaze at the camera	51%
Gaze at the right-page of the book	18%
Gaze at the left-page of the book	22%
Story related gaze	9%

Table 17: Amount of gaze of the embodied storyteller for the newsreader perspective

```
/**
 * Algorithm to determine the gaze point that will be selected next
 **/

rn = random(0,99) //select a random element from 100 elements
beginAt = 0;
FOR EACH gazePoint : availableGazePoints DO
   IF gazePoint.equals(C) AND beginAt <= rn < (beginAt + norm(amountOfGaze(C)))THEN
      nextGazePoint = C
      beginAt = beginAt + amountOfGaze(C)
   ELSE IF gazePoint.equals(S) AND beginAt <= rn < (beginAt + norm(amountOfGaze(S)))THEN
      nextGazePoint = S
      beginAt = beginAt + amountOfGaze(S)
   ELSE IF gazePoint.equals(LP) AND beginAt <= rn < (beginAt + norm(amountOfGaze(LP)))THEN
      nextGazePoint = LP
      beginAt = beginAt + amountOfGaze(LP)
   ELSE IF gazePoint.equals(RP) AND beginAt <= rn < (beginAt + norm(amountOfGaze(RP)))THEN
      nextGazePoint = RP
      beginAt = beginAt + amountOfGaze(RP)
   END IF
END FOR EACH
```

Algorithm 5: Select a gaze point from the available gaze points for newsreader perspective

Determine Gaze Duration

Finally the duration of the gaze for the newsreader perspective needs be determined. The newsreader annotations showed that the duration of the gaze at the camera is significantly longer (average duration was 10 seconds) than that the storyteller from "Elly en de wiebelwagen" gazes at the camera. The difference can partially be explained by the use of a teleprompter by the newsreader, so that he does not need to look at his papers.

Because of the large differences between gaze durations of the actual newsreader and actual storyteller, the gaze at the camera is adjusted. The assumption is made that when increasing the gaze duration at the camera, it is better suited for the newsreader perspective. The adjustment in the gaze behaviour also compensates the absence of the audience.

The gaze duration at the camera will be a multiple of the average found in the storyteller annotations. As a starting point three times the average gaze duration will be used (Table 16). During the implementation and evaluation phase of the gaze model it will be seen how this selected gaze duration at the camera will be evaluated.

The gazing points	**Mean duration (μ)**	**Standard deviation (σ)**
Gaze at the camera	4.125 (3*1.375)	0.6
Gaze at the right-page of the book	1.110	0.5
Gaze at the left-page of the book	1.074	1.0
Story related gaze	1.520	0.8

Table 18: Gaze durations derived from the storyteller annotations

```
/**
 * Interval of values that can be the duration of the gaze
 * Where μ and σ are the values for the selected gaze point found in table 18
 **/

Gaze duration(Gaze Point): [μ - σ, μ + σ] = {x ∈ (μ - σ) ≤ x ≤ (μ + σ)}
```

The set of values for the gaze duration of a specific gaze point for the newsreader perspective

5.3 Discussion

The elements from the gaze models discussed in chapter thee and the results from the data collected during the annotations are combined and used in the gaze model created for the Embodied Storyteller. The objective was to keep the gaze model context free, meaning no text or speech analysis is required to determine the different gaze behaviours during storyteller. In the end the gaze model is context free and consist out of three phases:

1. Determine the available gaze points for a certain time in the story
2. Select the next gaze point from the available gaze points
3. Determine the gaze duration.

Both viewing perspectives use the same gaze model, with the exception that they have their own specific gaze points, gaze frequencies and gaze durations. The gaze model enables the Embodied Storyteller to perform a different sequence of gaze behaviours every time it tells a story. The gaze model will be implemented in a Virtual Human in an existing framework, to see how the gaze model performs and if it actual displays varied gaze behaviours. The implementation and the used framework will be talked about in the next chapter.

6 IMPLEMENTATION

The two different viewing perspectives and the specific gaze models for the embodied storyteller for one of the viewing perspectives were implemented by using Elckerlyc [WR10]. Elckerlyc provides the ability to create an embodied storyteller in a virtual domain and select gaze behaviours during the embodied agent's storytelling. By using the possibilities of Elckerlyc a different sequence of gaze behaviours can be created every time a story is told.

6.1 Elckerlyc and BML

Creating an embodied agent can be a daunting task. Elckerlyc enables the user to manipulate an embodied agent in a virtual environment [WR10]. Elckerlyc uses the Behaviour Mark-up Language (BML) to enable the user to describe verbal and non-verbal behaviours on an abstract level. By using BML descriptions, gaze behaviours can be created that meet the terms of the gaze model for the embodied storyteller. BML is part of the SAIBA-framework [KK06]. The framework is the result of an international collaboration to create a multimodal framework for the verbal and non-verbal behaviours of an embodied conversational agent (ECA).

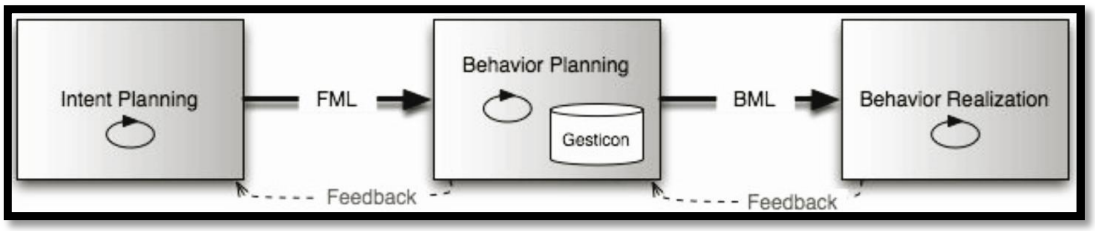

Figure 21: SAIBA framework for multimodal generation [KK06, p209]

The SAIBA-framework consists out of three stages. The first stage is planning the intent of the embodied agent, the second stage is planning the behaviour and the third and final stage is the realization of the behaviour of the embodied agent. (Figure 21). The connection between planning the intent and planning the behaviour of the embodied agent is done through the Functional Mark-up Language. The Functional Mark-up Language provides the semantic information relevant for planning the verbal and non-verbal behaviours. The specification of the Functional Mark-up Language is in the early stages of development and discussions towards a consensus in the used tags and their functions is ongoing [HK08]. The connection between planning of the behaviour and the realization of the behaviour is done through the Behavioural Mark-up Language. The Behavioural Mark-up Language provides a general description of the verbal and non-verbal behaviours that can be used to control the embodied storyteller. The different gaze behaviours of the embodied storyteller, as well as the realization of the speech, blinking and body movements will be discussed in the upcoming section.

6.2 The setting

Besides implementing the gaze model and the two viewing perspectives, time was spent on creating a compelling storytelling setting, similar to the setting found in " Elly en de wiebelwagen". The setting, the embodied agent, creating the viewing perspectives by placing the camera and the realization of speech, blinking and body-movements will be discussed in the upcoming sections.

Virtual Human

Elckerlyc is written in JAVA. A basic project serves as the starting point for the implementation of the gaze model and realization of the embodied storyteller. The first step is to select an Virtual Human that serves as the embodiment of the actual storyteller. Elckerlyc provides two basic Virtual Humans, one being a blue guy without any facial features and another Virtual Human called Armandia. Armandia (figure 22) has the possibility to blink with her eyes and move with her lips when speaking. The SAPI5 engine from Microsoft with Dutch speech synthesis was used to tell the story. Subsequently a setting is loaded into the virtual environment. The setting is a derivative of an already existing room. The original room called "the psychoroom", created by Solano et al. [MR11], serves as the background for a psychologist. The placement of the objects, as well as the Virtual Human are repositioned to a situation in which Armandia sits in a chair in front of a book case (figure 22 and 23). The final element loaded into the environment is the book (http://sketchup.google.com/), it is a representation of the relative large book as found in "Elly en de wiebelwagen".

Placement of the camera

Figure 22 and 23 are the two resulting positions for the camera, the first representing the frontal view for the newsreader perspective. The other a perspective with a sideways angle, representing the viewing perspective as found in "Elly en de wiebelwagen". The placement of the camera was done through trial,- and error. Small alterations in the camera angle, as well as the positioning on the x, y, z axis resulted in a positioning aimed towards the newsreader and storyteller perspective.

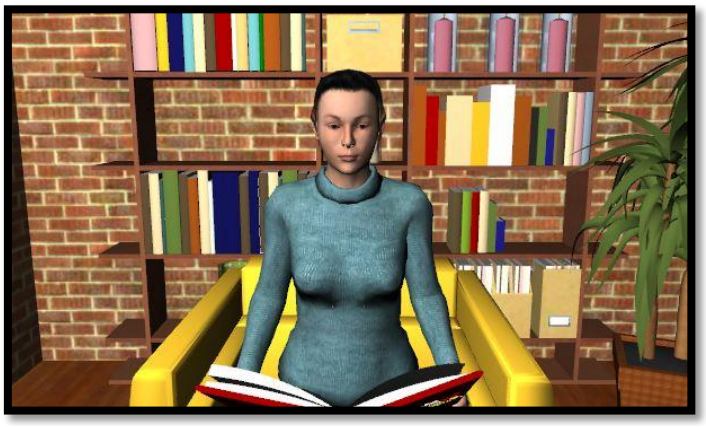

Figure 22: The setting, the book and Armandia from the newsreader perspective

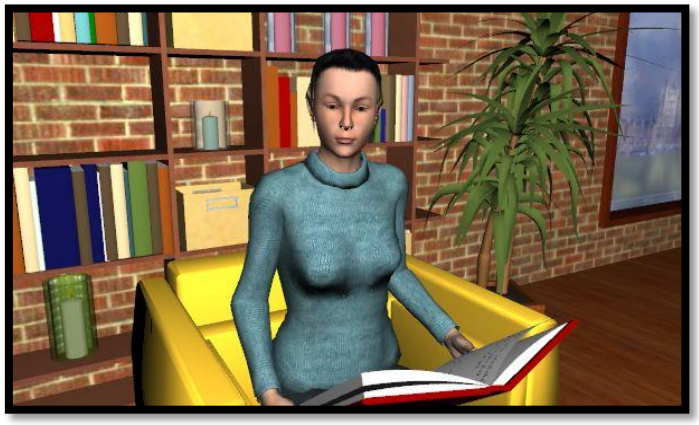

Figure 23: The setting, the book and Armandia from the storyteller perspective

Speech, blinking and body movements

To make the agent act more human-like, blinking and a simple body movement was added to the embodied storyteller. Including the speech behaviour of Armandia, beneath you can see a description of the first three behaviours written in BML. Both the blinking emitter and the movements of a joint are extensions on the BML description developed by the Human Media Interaction group at the University Twente. The extensions can be recognized by the *<bmlt/>* tag, which stands for BML Twente.

The first behaviour is the blinking emitter, which enables Armandia to blink with her eyes at a specific time. The blinking starts at time zero and has a range of half a second with an average waiting time of four seconds. The *avgwaitingtime* is the time between different blinks. The *range* indicates the maximum variation in waiting time for the next blink. To add a basic body movement to Armandia a Perlin noise was added to one of the joints. The joint (*vt6*) in the lower back of Armandia, provided the best results to simulate a simple body movement. The movement is enabled by setting a *basefreqx* and the *baseamplitudex* for the movement of the joint over the x-axis. The joint movement starts at time 0 and finishes after 30 seconds. The Perlin noise is a function that creates a smooth transition between the different points of the movement.

Finally the speech tag is added to the BML description. The story is told in Dutch and is the first part of the fairy tale Snow White. The default reading speed of Armandia was a little too fast and the *rate* tag was added to the BML description to lower the absolute reading speed (*absspeed*).

```
<bml id="story" xmlns:bmlt="http://hmi.ewi.utwente.nl/bmlt">
 <bmlt:blinkemitter id="blinkemitter1" start="0" range="0.5" avgwaitingtime="4"/>
   <bmlt:noise id="noisevt6" type="perlin" joint="vt6" start="0" end="30">
     <bmlt:parameter name="basefreqx" value="0.6"/>
     <bmlt:parameter name="baseamplitudex" value="0.10"/>
   </bmlt:noise>
   <speech id="speech1" start="4">
   <description priority="1" type="application/msapi+xml">
    <sapi>
     <speak>
      <rate absspeed="-7">
        De vorstin prikt zich aan de naald en er vallen drie druppels bloed in de sneeuw....
      </rate>
     </speak>
    </sapi>
   </description>
  </speech>
  <!-- Next the gaze behaviours are created -->
</bml>
```

Example of Behaviour Markup language extended with BMLTwente tags (For a description of the BMLTwente extensions see: http://elckerlyc.ewi.utwente.nl/wiki/BMLT#BMLTwente)

6.3 The gaze behaviours

After creating the BML behaviours for the blinking of the eyes, the body movements and the speech of Armandia the gaze behaviours are created. From the gaze model in chapter five it is seen that the selection of a gaze behaviour is split into three parts:

- Part I: Depending on the current conditions determine the available gaze points that can be selected as the next gaze behaviour.
- Part II: Create a probability distribution for the available gaze points and the likelihood that a certain gaze point will be selected.
- Part III: Provide the duration of the selected gaze point that will occur next.

The different parts of the gaze model have been implemented in JAVA (descriptions of the algorithms for the different part of the gaze model can be found in chapter five) and are combined with the BML-description of a gaze behaviour. The specific probabilities and gaze durations derived from the annotations can be found in chapter four. In total there are four gaze behaviours for the storyteller perspective and three for the newsreader perspective. The general construction of a gaze behaviour in BML is:

```
<bml>
    <gaze id="" start="" ready="" relax="" end="" type="" modality="" target="" dynamic=""/>
</bml>
```

Example of a general description of a gaze tag in the Behaviour Mark-up Language

Attribute	
id	Unique ID that allows referencing to a particular <bml/> behaviour. ID 'bml' is reserved.
target	A reference towards a target instance that represents the target direction of the gaze.
dynamic	Indicates that the position of the target in the environment is dynamic
type	The type of gaze behaviour (AT..)
modality	The modality (head, neck, torso) used when gazing, i.e. the NECK is used when gazing
start	The gaze starts to move to new target
ready	The gaze target is acquired
relax	The gaze starts to return to is default direction
end	The gaze returned to default direction

Table 19: The sync attributes and their meaning (For a detailed description of BML and the available attributes see http://www.mindmakers.org/projects/bml-1-0/wiki/Wiki)

Because a new gaze target is selected after each gaze, there is no default return state for the end of gaze. Therefore the relax and end attribute have been omitted from the gaze specification for the gaze behaviours of the embodied storyteller. Figures 24 till 39 show the different gaze behaviours accompanied with the BML description for both the storyteller and newsreader perspective.

Storyteller perspective

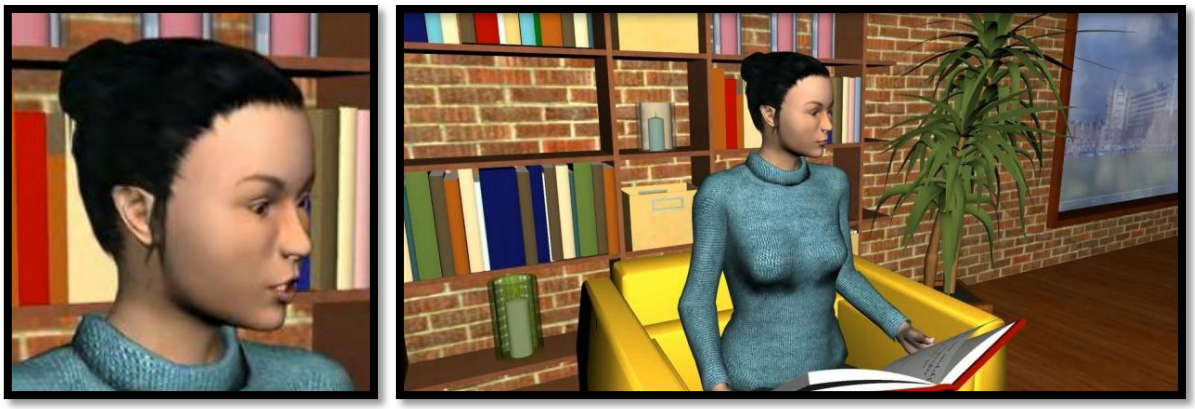

Figure 24: Gaze at the left-page of the book

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="topleftbook"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the left-page of the book

Figure 25: Gaze at the audience

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="centerofaudience"
       dynamic="false"/>
</bml>
```

BML Description of the gaze at the audience

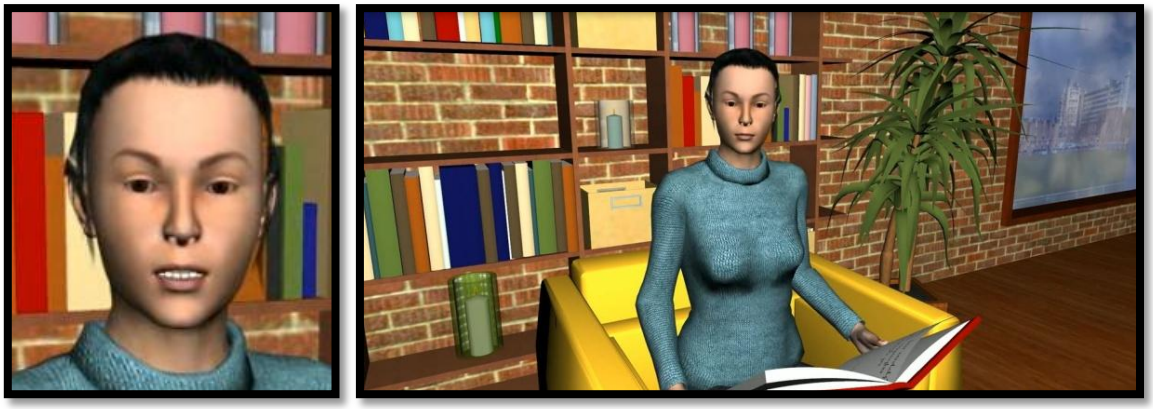

Figure 26: Gaze at the camera

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="camera"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the camera

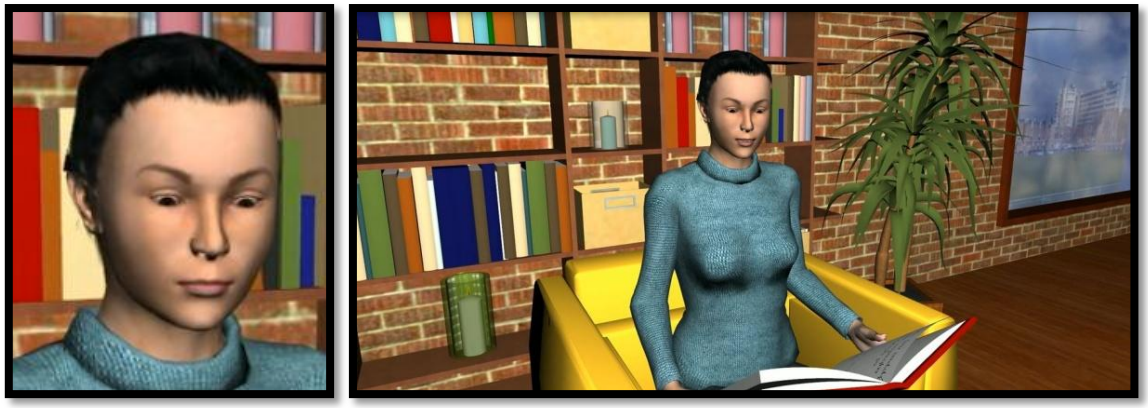

Figure 27: Gaze at the right-page of the book

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="toprightbook"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the right-page of the book

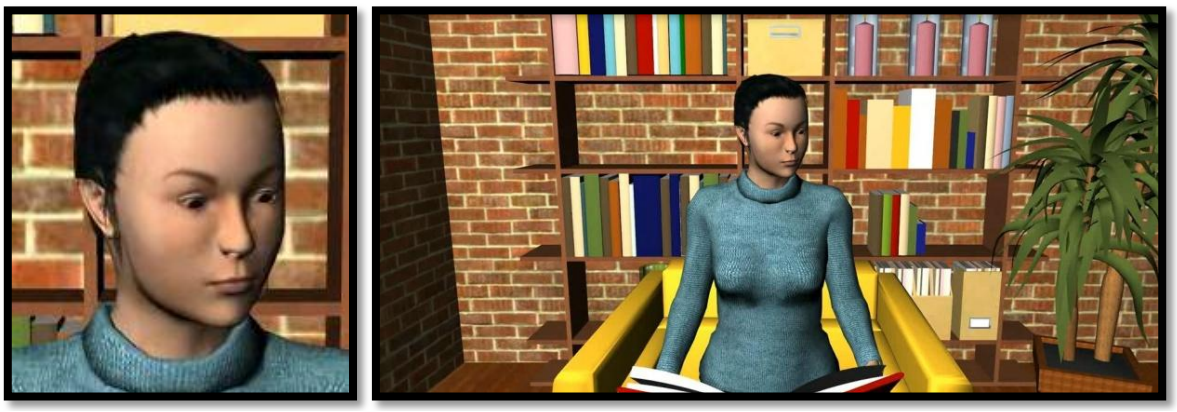

Figure 28: Gaze at the left-page of the book

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="topleftbook"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the left-page of the book

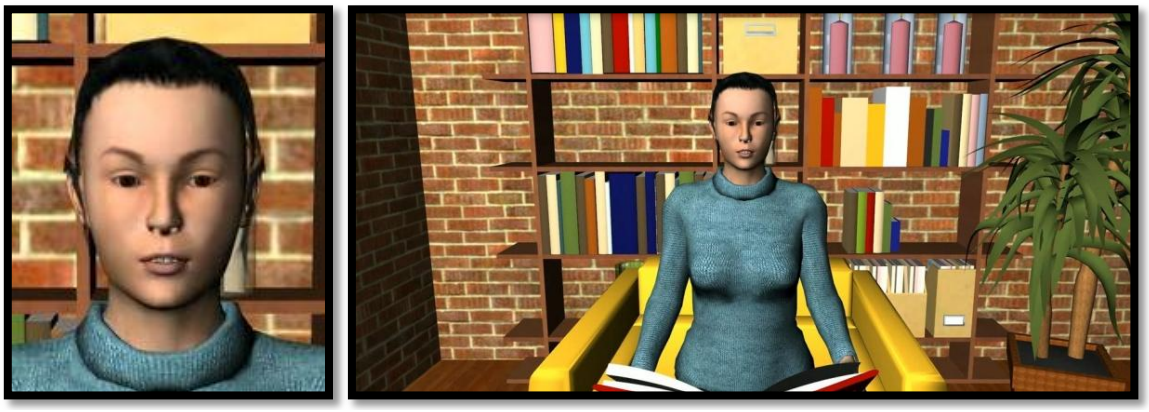

Figure 29: Gaze at the camera

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="centre"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the camera

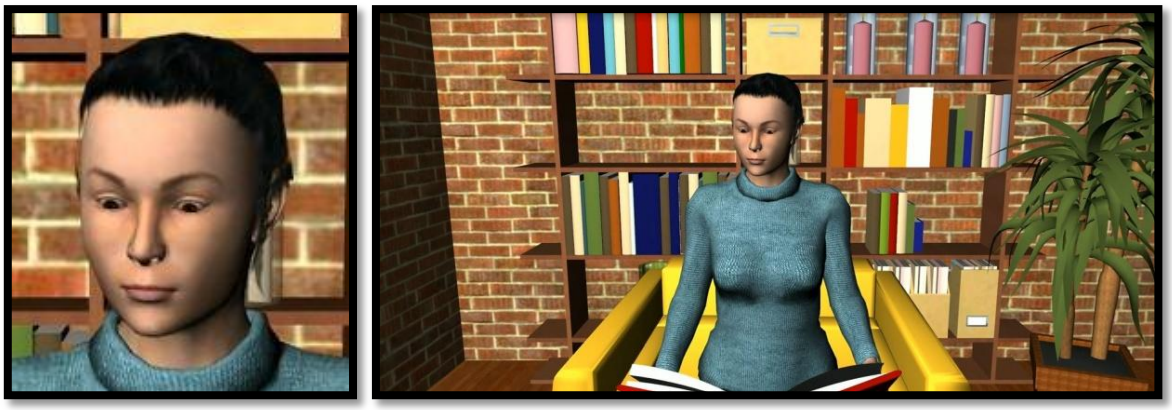

Figure 30: Gaze at the right-page of the book

```
<bml>
    <gaze id="gaze1" start="" ready="" type="AT" modality="NECK" target="toprightbook"
        dynamic="false"/>
</bml>
```

BML Description of the gaze at the right-page of the book

6.4 Discussion

Elckerlyc provided the ability to use an existing Virtual Human (Armandia) as an embodied storyteller. The embodied storyteller has the possibility to use gaze behaviours to augment the storytelling. The gaze behaviours are determined by the gaze model, as described in chapter seven. However due to the lack of readily available documentation and the early development stage of Elckerlyc, quite some time was spend on the creation of a compelling setting with an embodied storyteller that gazed fluently.

One limitation of Elckerlyc is the maximum angle of Armandia's gaze. During the implementation of the gaze model it was the idea to enable Armandia to read from top to the bottom of a page. Due to this limitation it was only possible to gaze at the top of the book. Subsequently gaze behaviours are usually a combination of the movement of the eyes and the movement of the neck. These modalities don't always move at the same time. However with the gaze descriptions in the Behaviour Mark-up Language, the modalities did always move at the same. Optimally one would separate these modalities to create a more human-like gaze behaviour.

Story-related gazing was excluded from the implementation, because it requires context from the story told and a specific gaze behaviour designed for the current context. It would require an added amount of time to specifically design one gaze behaviour for one story-related gaze. Compared to the reusable gaze behaviours, like the gaze at the book, camera or audience, a story-related gaze is too time consuming to implement. Head movements were also tested as possible behaviours of Armandia. As mentioned in an earlier section head movements often accompany gaze behaviour and a simple nod or shake (yes/no) should be easy to implement. However the nods and shakes appeared unnatural when used by Armandia and not beneficiary to the entire viewing experience and were omitted from the final implementation.

Armandia, the setting, the verbal and non-verbal behaviours (gazing, body movement) and the two perspectives create two entire different viewing experiences, which was the objective in the first place. Due to the implementation of the gaze model, Armandia has the possibility to selected different gaze behaviours with different gaze durations and gaze frequencies. Which can result in a different viewing experience every time a story is told. In the upcoming section the setting, the gaze model, the behaviours and the two perspectives will be tested by creating two video fragments, one for each of the viewing perspectives. By using a user survey the general opinion of the user will be acquired. In the end it will be seen how the gaze behaviours are rated and if there is any preference for one the two viewing perspectives.

7 EVALUATION

To reiterate, the main research question is to see if there is any preference for a certain viewing perspective. In which gazing behaviours and head movements are the main non-verbal behaviours. By exploring existing literature and annotating a real-world storyteller and newsreader, the duration and the frequency of the gaze behaviours and head movements were determined. This information was used to formulate a gazing and head movement model. The gaze model was implemented in an embodied storyteller within Elckerlyc.

7.1 Experimental setup

A short user survey was carried out to evaluate the different gaze behaviours of the embodied storyteller. Two video fragments were created. One video fragment showed the embodied agent from the newsreader perspective. The other video fragment showed the embodied agent from the storytelling perspective. In both video fragments the embodied storyteller tells the first part of fairytale Snow White. Text-To-Speech was used to tell the story. The average length of both video sessions was around 44 seconds. In this time all of the gaze behaviours are performed multiple times. To ensure that the order in which the video fragments were seen, didn't influence the assessment of the viewing perspectives and gazing behaviour, two versions of the questionnaire were created. The only difference being the order in which the video fragments were seen. Participants could view each of the videos fragments as much as they wanted.

In total the questionnaire consisted of five general questions regarding gender, age, educational level, computer skills and familiarity with ECAs and twenty-four questions regarding the video fragments. After each video fragment questions followed about the gaze duration and the gaze frequency of a specific gaze point. Participants were also asked to provide their general opinion about the displayed gaze behaviours, the posture and accompanying body movement (table 20). After the video fragment with the storytelling perspective questions were also asked about the presence of the audience and the gaze behaviours towards the audience (table 21). The questionnaire ended with the question if the participant had a preference for one of the fragments, if they had to view the total story of Snow White (table 22). Participants were contacted by email and through social media (Twitter and Facebook). Always keeping in mind that there would be an equal division between the two questionnaires.

Frequency of the gaze at the book	Do you think the storyteller looks often enough at the book? *1=far too few, 2=too few, 3=good, 4=too often and 5=far too often*
Duration of the gaze at the book	At the moment the storyteller looks at the book, what do you think about the duration she looks at the book? *1=far too short, 2=too short, 3=good, 4=too long and 5=far too long*
Frequency of gaze at the camera	Do you think the storyteller looks often enough at the camera? *1=far too few, 2=too few, 3=good, 4=too often and 5=far too often*
Duration of the gaze at the camera	At the moment the storyteller looks at the camera, what do you think about the duration she looks at the camera? *1=far too short, 2=too short, 3=good, 4=too long and 5=far too long*
General opinion about the gaze	What do you think in general about the gaze behaviours of the storyteller? *1=very bad, 2=bad, 3=neutral, 4=good, 5= very good*
General opinion about the posture and body movements	What do you think in general about the posture of the storyteller and the accompanying body movements? *1=very bad, 2=bad, 3=neutral, 4=good, 5= very good*

Table 20: Questions for both viewing perspectives

Presence of the audience	Do you have the feeling that besides you and the storyteller, there are other people (audience) in the same room facing the storyteller? *1=yes, 2=no*
Frequency of the gaze at the audience	Do you think the storyteller looks often enough at the audience? *1=far too few, 2=too few, 3=good, 4=too often and 5=far too often*
Duration of the gaze at the audience	At the moment the storyteller looks at the audience, what do you think about the duration she looks at the audience? *1=far too short, 2=too short, 3=good, 4=too long and 5=far too long*

Table 21: Questions regarding the audience in the storyteller perspective

Preference	Which video fragment would you prefer if you had to view the full story of Snow White? 1=video fragment 1, 2=video fragment 2, 3=no preference

Table 22: Question regarding preference for a certain viewing perspective

7.2 Results

Mirror, mirror upon the wall, Who is the fairest fair of us all? It answered O Lady Queen, though fair ye be, Snow-White is fairer far to see.

Participants

Thirty-four participants completed one of the two questionnaires. Nineteen participants completed questionnaire 1 (table 23) in which the storyteller perspective was shown first, followed by the newsreader perspective. Fifteen participants completed the questionnaire 2 in which the newsreader perspective was shown first, followed by the storyteller perspective (table 23). As mentioned before the participants were contacted by e-mail and social media. Participants were told that they were going to see two video fragments in which they would see small portions of Snow White. They were asked to view both video fragments and asked to answer the questions as honest as possible. They were also told that the story was told using text-to-speech and they should not let them be distracted by the use of the text-to-speech.

Questions	**Options**	**Questionnaire 1 (n=19)**	**Questionnaire 2 (n=15)**
Gender	Male: Female:	10 (53%) 9 (47%)	8 (53%) 7 (47%)
Age	Younger than 25: 25-35: 36-45: 46-55: 56-65: Older than 65:	3 (16%) 7 (37%) 1 (5%) 5 (26%) 3 (16%) 0 (0%)	4 (27%) 8 (53%) 1 (7%) 1 (7%) 1 (7%) 0 (0%)
Educational level	Primary educational level: High school: Lower vocational (LBO): Secondary vocational (MBO): Higher vocational (HBO): Scientific education (WO):	1 (5%) 0 (0%) 1 (5%) 6 (32%) 5 (26%) 6 (32%)	0 (0%) 0 (0%) 0 (0%) 0 (0%) 9 (60%) 6 (40%)
Computer skill	Very bad: Bad: Neutral: Good: Very good:	0 (0%) 0 (0%) 8 (42%) 9 (47%) 2 (11%)	0 (0%) 0 (0%) 2 (13%) 10 (67%) 3 (20%)
Familiarity with ECAs	Has never heard about ECAs: Has heard of the term ECA: Knows ECAs, has not worked with them: Knows ECAs, worked with them:	15 (79%) 3 (16%) 0 (0%) 1 (5%)	12 (80%) 2 (13%) 1 (7%) 0 (0%)

Table 23: Demographic results for both questionnaires, with absolutes numbers and percentages (in parenthesis)

Duration and the frequency of the gaze

The average rating of the questions regarding the duration of the gaze and the gaze frequency can be found in tables 24 and 26. Distinction is made between the two different viewing perspectives and if the viewing perspective was shown first or last. Table 25 shows the results for the presence of the audience for the storyteller perspective and indicate that most people acknowledged the presence of an audience. Regarding the presence of the audience, one participant failed to provide an answer.

The results of the gaze duration and gaze frequency are also visualized in a quadrant. The quadrant shows a the spread between the different ratings for both viewing perspectives. Also it is seen if the moment of when the video fragment was seen (first or last) influences the rating provided by the participant. In total four quadrants were created that compare the duration of the gaze and frequency of the gaze for each of the different gaze points (book, camera, audience):

- Figure 31 and table 27: Provides the rating (x=duration, y=frequency) of the different gaze points (book, camera and audience) of the storyteller perspective when viewed as the first or last video fragment.
- Figure 32 and table 28: Provides the rating (x=duration, y=frequency) of the different gaze points (book and the camera) of the newsreader perspective when viewed as the first or last or last video fragment.
- Figure 33 and table 29: Provides the rating (x=duration, y=frequency) of the different gaze points (book, camera and audience) of the storyteller perspective and newsreader perspective when viewed as the first video fragment.
- Figure 34 and table 30: Provides the rating (x=duration, y=frequency) of the different gaze points (book, camera and audience) of the storyteller perspective and newsreader perspective when viewed as the second video fragment.

The significance of each of the provided ratings is calculated by means of a T-TEST. Values lower than 0.05 are considered statistically relevant and only the relevant values will be mentioned. The T-TEST is a two-tailed test between two distributions, both with their own variance.

Storyteller (average result)	**Viewed as first video fragment**	**Viewed as second video fragment**
Frequency of gaze at the book	2.26 (0.65)	2.87 (0.83)
Duration of the gaze at the book	2.47 (0.70)	2.67 (0.82)
Frequency of gaze at the camera	2.58 (0.77)	2.73 (1.03)
Duration of the gaze at the camera	2.74 (0.56)	2.67 (0.82)
Frequency of gaze at the audience	3.21 (0.85)	3.07 (0.80)
Duration of the gaze at the audience	3.05 (0.78)	3.00 (1.07)

Table 24: Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for the storytelling perspective.

	Options	**Viewed as first video fragment**	**Viewed as second video fragment**
Presence of the audience	YES NO	16 (84%) 3 (16%)	10 (67%) 4 (27%)

Table 25: Presence of the audience in the storytelling perspective

Newsreader (average result)	**Viewed as first video fragment**	**Viewed as second video fragment**
Frequency of gaze at the book	2.27 (1.10)	2.37 (0.96)
Duration of the gaze at the book	2.40 (0.99)	2.37 (0.68)
Frequency of gaze at the camera	3.40 (0.99)	3.05 (0.91)
Duration of the gaze at the camera	3.20 (1.15)	3.11 (0.81)

Table 26: Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for the newsreader perspective.

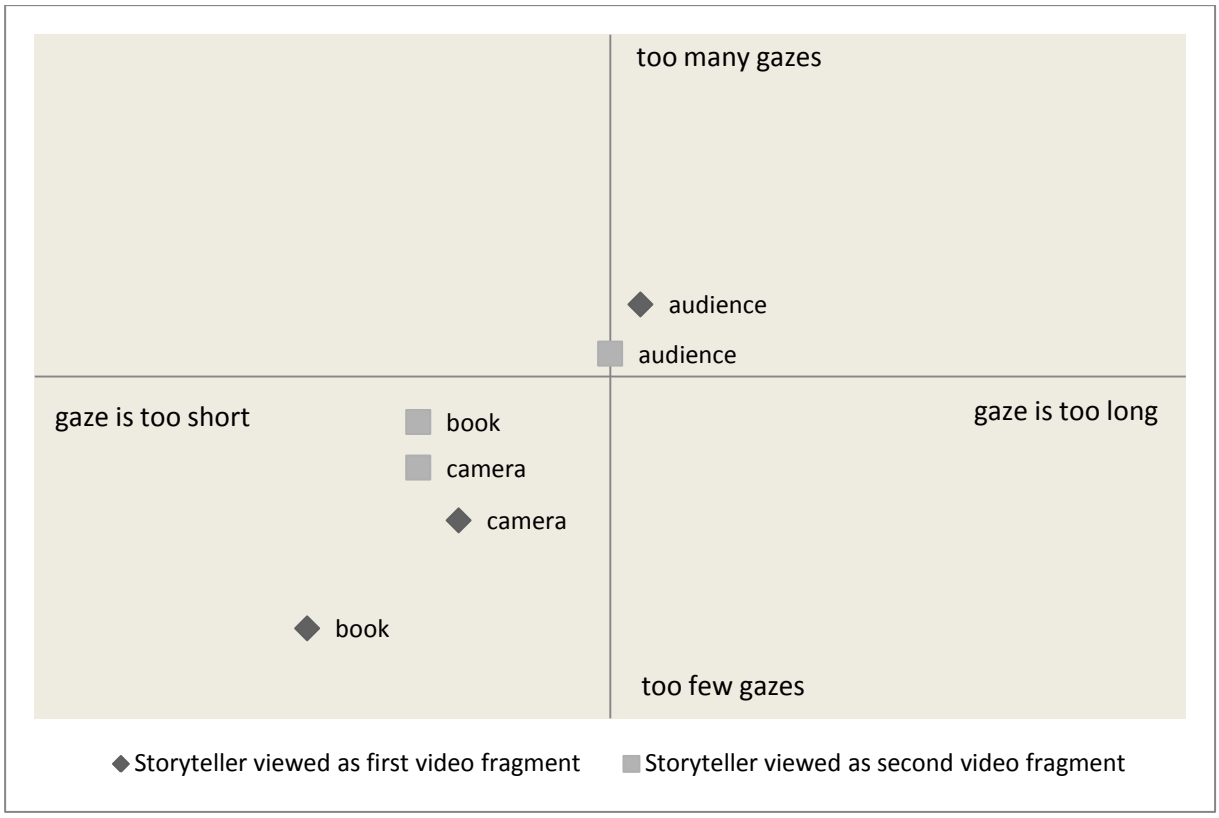

Figure 31: Comparison between the sequence of the viewing of the storytelling perspective

Storyteller	**Viewed as first video fragment**		**Viewed as second video fragment**	
	Gaze frequency	**Gaze duration**	**Gaze frequency**	**Gaze duration**
Gaze at the book	2.26 (0.65)*	2.47 (0.70)	2.87 (0.83)*	2.67 (0.82)
Gaze at the camera	2.58 (0.77)	2.74 (0.56)	2.73 (1.03)	2.67 (0.82)
Gaze at the audience	3.21 (0.85)	3.05 (0.78)	3.07 (0.80)	3.00 (1.07)

Table 27: Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for the storytelling perspective. * are the values that are significantly different.

The axes in figure 31 represent the ratings for the duration of the gaze (x-axis) and the frequency of the gaze (y-axis). The centre of both axes intersect at the point where the participants rated the gaze behaviour as good (which is valued at 3, see table 27 for ratings for the duration of the gaze and the frequency of the gaze).

The quadrant shows no large differences in the results for the gaze at the camera and the gaze at the audience. This is also indicated by their p-values. However for the gaze at the book there is a large difference between the rating for the gaze frequency when seen first and seen second. A significant difference was found with a p-value of 0.03.

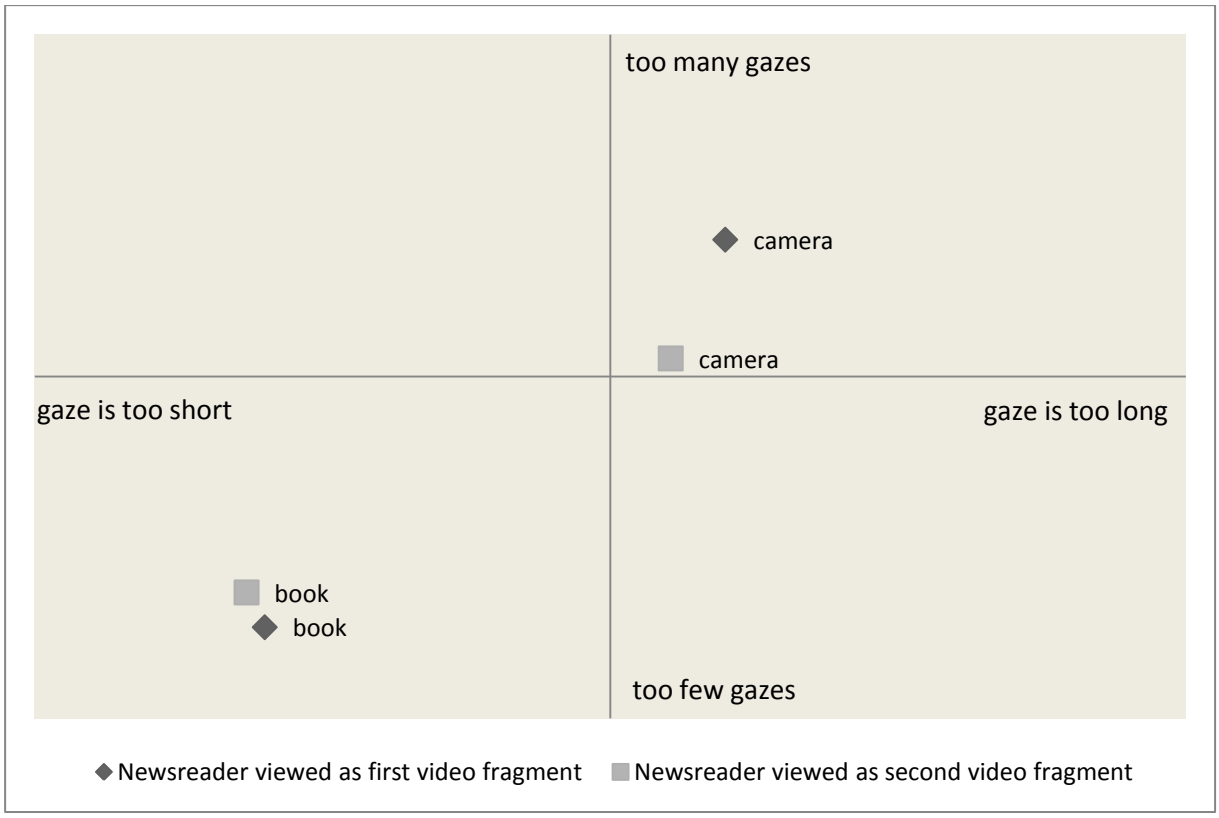

Figure 32: Gaze length and gaze amount for the newsreader-perspective

Newsreader	Viewed as first video fragment		Viewed as second video fragment	
	Gaze frequency	Gaze duration	Gaze frequency	Gaze duration
Gaze at the book	2.27 (1.10)	2.40 (0.99)	2.37 (0.96)	2.37 (0.68)
Gaze at the camera	3.40 (0.99)	3.20 (1.15)	3.05 (0.91)	3.11 (0.81)

Table 28: : Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for the newsreader perspective.

The axes in figure 32 represent the ratings for the duration of the gaze (x-axis) and the frequency of the gaze (y-axis) for the newsreader perspective when viewed as the first and last video fragment. The centre of both axes intersect at the point where the participants rated the gaze behaviour as good for both the frequency and the duration (which is valued at 3, see table 28 for ratings).

The quadrant shows no large difference in the results for the gaze at the book. For the gaze at the camera it is seen that they values show a reasonable difference for the frequency of the gaze. When the newsreader is viewed as the first video fragment the gaze at the camera is valued as to frequently. However no statically relevant value is found and the comparison of the frequency has a p-value of 0.30.

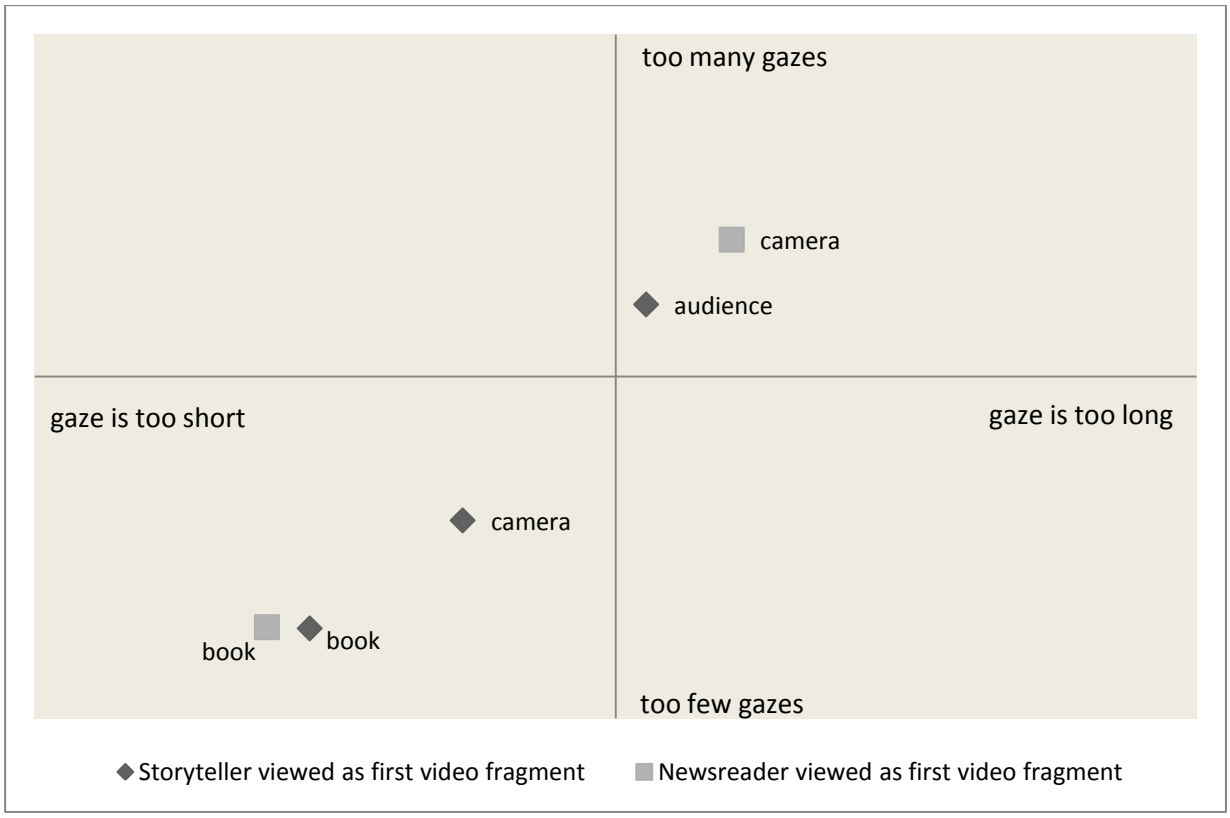

Figure 33: storyteller versus newsreader perspective both viewed as the first video fragment

Perspective viewed as first video fragment	Storyteller		Newsreader	
	Gaze frequency	Gaze duration	Gaze frequency	Gaze duration
Gaze at the book	2.26 (0.65)	2.47 (0.70)	2.27 (1.10)	2.40 (0.99)
Gaze at the camera	2.58 (0.77)*	2.74 (0.56)	3.40 (0.99)*	3.20 (1.15)
Gaze at the audience	3.21 (0.85)	3.05 (0.78)		

Table 29: Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for both perspective when seen as the first video fragment. * indicate significantly different values.

The axes in figure 33 represent the ratings for the duration of the gaze (x-axis) and the frequency of the gaze (y-axis) for the newsreader perspective when viewed as the first and last video fragment. The centre of both axes intersect at the point where the participants rated the gaze behaviour as good for both the frequency and the duration (which is valued at 3, see table 29 for ratings).

The quadrant shows that the gaze at the book is rated for both the newsreader and the storyteller perspective as too short and too infrequent. The gaze at the camera shows a large difference and is statistically relevant with a p-value of 0.01 for the frequency of the gaze.

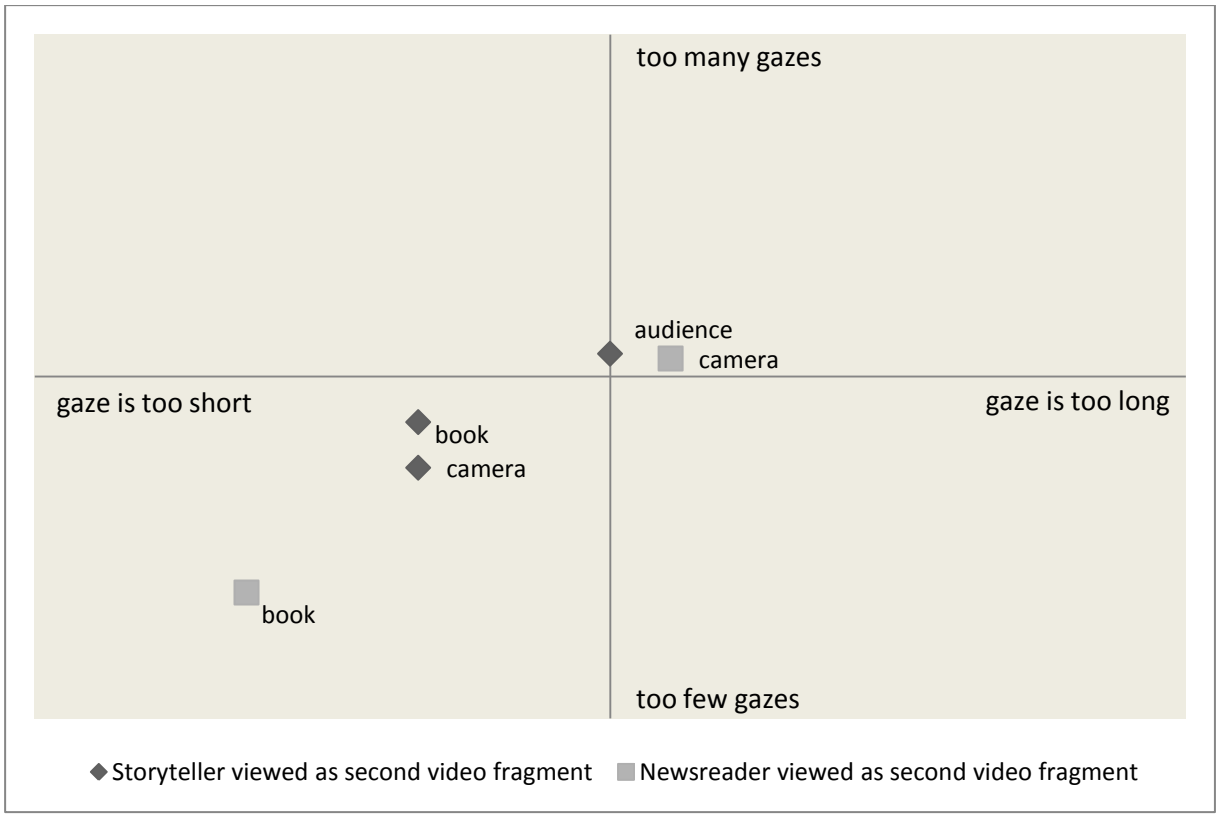

Figure 34: storyteller versus newsreader both viewed as the second video fragment

Perspective viewed as second video fragment	Storyteller		Newsreader	
	Gaze frequency	Gaze duration	Gaze frequency	Gaze duration
Gaze at the book	2.87 (0.83)	2.67 (0.82)	2.37 (0.96)	2.37 (0.68)
Gaze at the camera	2.73 (1.03)	2.67 (0.82)	3.05 (0.91)	3.11 (0.81)
Gaze at the audience	3.07 (0.80)	3.00 (1.07)		

Table 30: Mean and the standard deviation (in parentheses) for the gaze frequency and gaze duration for both perspective when seen as the second video fragment.

The axes in figure 34 represent the ratings for the duration of the gaze (x-axis) and the frequency of the gaze (y-axis) for the newsreader perspective when viewed as the first and last video fragment. The centre of both axes intersect at the point where the participants rated the gaze behaviour as good for both the frequency and the duration (which is valued at 3, see table 30 for ratings).

The quadrant shows that newsreader and the storyteller are valued very differently, in which the gaze at the book for the newsreader perspective is too infrequent. However none of the values are statically relevant value with the lowest p-value being 0.11 with the frequency for the gaze at the book.

General opinions

After the questions regarding the gaze duration and the frequency of the gaze, participants were asked to provide their opinion about the gaze behaviour of the embodied storyteller in general and posture of the storyteller with the accompanying body movements. The results of the questions can be found in table 31 and 32. The participants were also asked to clarify their chosen answers. Some of the provided answers are discussed in the upcoming discussion section at the end of this chapter. Figure 35 and Figure 36 visualize the average rating for each of the viewing perspectives and the moment in which the video fragment was seen (first or last).

The results show a rating for the general gaze opinion and the posture with the body movement less than neutral. For the rating of the general gaze of the newsreader, when seen as first video fragment, is rated the lowest with 2.20. The storyteller, when seen as the first video fragment, is rated the highest with 2.79. For the rating of the body movements the newsreader, when seen as second video fragment, is rated the lowest. The storyteller, when seen as the second video fragment, has the highest rating with 3.00. None of the results are statistically relevant, with p-values ranging from 0.07 to 0.92.

The comparison between of the general opinion of the gaze of the storyteller and newsreader perspective (figure 36) when seen as the first video fragment has the p-value of 0.07, which is still too high to be relevant. It can suggest a slight preference for the storytelling perspective when looking at the general gaze opinion.

Storyteller	**Viewed as first video fragment**	**Viewed as second video fragment**
The displayed gaze behaviour	2.79 (0.98)	2.60 (0.91)
The posture and body movements	2.84 (1.01)	3.00 (0.58)

Table 31: Gaze and the posture movements of the agent in the storytelling perspective

Newsreader	**Viewed as first video fragment**	**Viewed as second video fragment**
The displayed gaze behaviour	2.20 (0.86)	2.63 (0.83)
The posture and body movements	2.93 (0.88)	2.58 (0.90)

Table 32: Gaze and the posture movements of the agent in the newsreader perspective

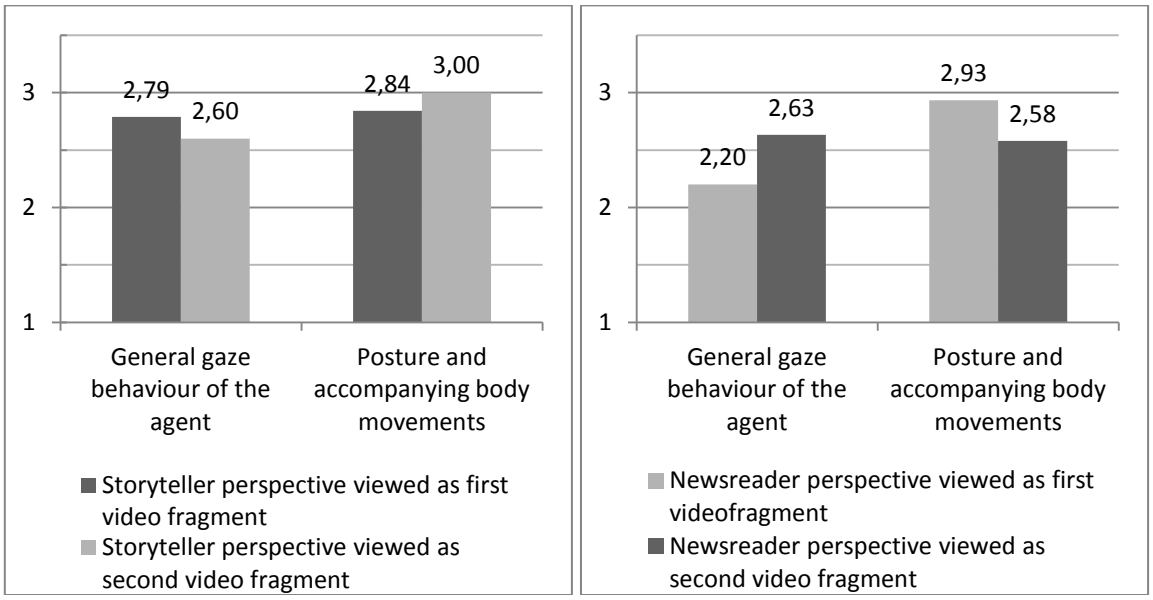

Figure 35: Comparing the sequence of the video fragments for both perspectives

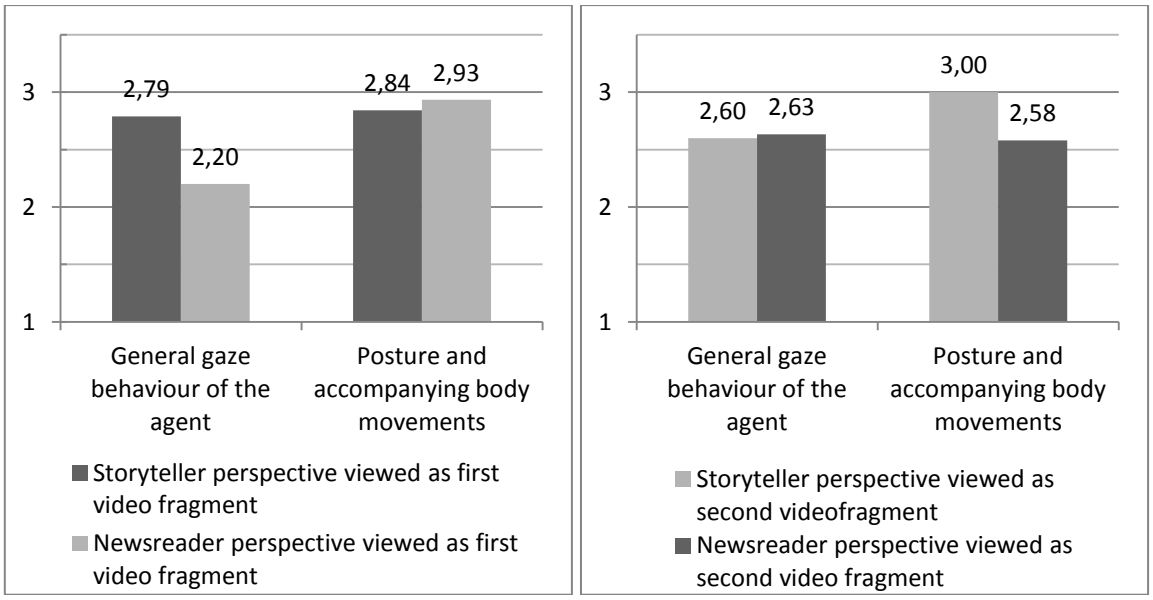

Figure 36: Comparing both perspectives for the different sequences

Preference

Finally participants were are asked to provide their preference for one of the two video fragments when viewing the complete story of Snow White. This would be either the first or second video fragment they had seen. There was no specific reference made to the viewing perspective. The results from table 23 show a slight preference for the storytelling perspective when shown as the last fragment. Participants were also asked to clarify their answer. Some of the provided answers will be used in the discussion section at the end of this chapter.

	Options	**Questionnaire were storyteller perspective was seen first**	**Questionnaire were newsreader perspective was seen first**
Preference	Video fragment with storyteller perspective	9 (47%)	9 (60%)
	Video fragment with newsreader perspective	9 (47%)	5 (33%)
	No preference	1 (5%)	1 (7%)

Table 33: Preference for video fragment

7.3 Discussion

Overall the participants rated none of the gaze behaviours as clearly negative. Some of the gazes were valued as too short and too few (book, camera for the storyteller perspective and the book for the newsreader perspective). Other gaze behaviours were rated as too often and too long (gaze at the audience for the storyteller perspective and the gaze at the camera for the newsreader perspective). The value provided for the gaze at the book can partially be explained by the expectations of the participants. Participant remarked that they could not believe that the embodied storyteller knew the complete story by heart and they expected that the embodied storyteller should gaze at the book more often and longer.

When reading through all the other individual comments regarding the gaze behaviours, the posture and body movements it is seen that they all could use some improvements. One participant stated that the embodied storyteller gazed at an audience at the lower right corner when viewing the second video fragment from the newsreader perspective (see figure 28, chapter six). However this was the embodied storyteller gazing at the right-page of the book. The particular participant had the expectation that there would be an audience in the newsreader perspective, because of the earlier experience with the storyteller perspective. For the storyteller perspective questions were asked regarding the existence of an audience, participants commented that the embodied storyteller gazed at the floor, gazed underneath the camera or that the gaze of the storyteller made them feel uneasy.

Poppe et al. [PR07] researched user's accuracy in correctly guessing the current gaze of an avatar. The experiment consisted out of an avatar positioned in a virtual meeting room. Opposite the avatar ten balls were placed at a distance of 1.5 meters at eye height. Different viewpoint for the user were tested and during the experiment users were asked to state at which ball the agents head currently was directed. They found that a higher viewpoint allowed for a better discrimination of the balls, which yielded a lower identification-error score. For the Embodied Storyteller this could mean that the errors regarding the current gaze of the Embodied Storyteller, might be due to the current position of the camera for the viewing perspectives. Subsequently there are the evaluations for the posture and body movements of the Embodied Storyteller. Participants remarked that they found the Embodied Storyteller to be static, wooden like and an they noticed an overall absence of emotions.

Did the sequence of the viewed video fragments influence the ratings of the participants? From the opinions participants provided one should say yes. Reading the comments of the participants one can see that the second video fragment is compared to the first video fragment. The data only shows two statistically relevant values. The first value is the found when comparing the storyteller perspective, either seen as first or last video fragment. This comparison shows a statically relevant value for the comparison of the gaze frequency at the book (p-value of 0.03). The next value is seen when comparing the two viewing perspectives with each other, when seen as the first video fragment. This comparison show a p-value of 0.01 for the frequency of the gaze at the camera. Because the sequence of the seen video fragments is the only apparent difference between the two user surveys, one would say that this is determining factor for the significant values. However the other comparisons between the sequence of seen video fragments show no significant values.

Regarding the final question about the users preference for either of the two video fragments, one mistake was made. The video fragments should have been added to the final question. This to ensure that people could once again see the video fragments when forming an opinion about their preference. From the results it is seen that when the storyteller perspective is seen first, there is no clear preference for one or the other. 47% of the participants preferred the storyteller perspective, 47% of the participants preferred the newsreader perspective and one participant (5%) had no preference. On the other hand when the storyteller perspective is seen last, there is a clear preference, with 60% of the participants preferring the storyteller vs. 33% preferring the newsreader perspective and one participant (7%) had no preference.

"Not a lot of emotion"
"The gaze behaviours carry the story"
"Varied gaze behaviour, not static"
"She says too much text without looking at the book. Does not look realistic"

Quotes regarding the gaze of Armandia (for the original Dutch quotes see CD)

"I saw more clearly that she was reading from the book and the shift between reading and looking up appeared more natural" [Preferred the newsreader perspective]
"The first video takes you into the story." [Preferred the storyteller perspective]

Quotes regarding the preference when storyteller perspective was seen first

"In the first video clip, the narrator looks at the camera more often. Because of this I feel more involved in the story." [Preferred the newsreader perspective]
"In the second clip, I had more the idea that children were sitting on the floor listening." [Preferred the storyteller perspective]

Quotes regarding the preference when newsreader perspective was seen first

8 CONCLUSIONS

Embodied Conversational Agents (ECA) have come a long way in the recent decade. Not only in their visual representation, but also in their non-verbal and verbal abilities. The applications of the ECAs range from educational, informational or have the possibility to assist us with certain tasks. Different projects at the Human Media Interaction group (part of the Electrical Engineering, Mathematics and Computer Science Department at the University of Twente) are aimed at using ECAs for one of the above mentioned activities.

Take for example the Virtual Storyteller [ST08], which has the ability to generate and tell the generated stories automatically. The non-verbal abilities of the narrator in the Virtual Storyteller [ST08] is limited to behaviours that consist of a look-at and look-away state for the narrator, partly because only the top part of the narrator is visible in the virtual environment. How can this viewing experience be improved? Not only improvement can be made to the gaze behaviours, but also the viewing perspective for the user plays a role in the viewing experience. This led to a research question that asks which of two viewing perspectives provides the best viewing experience? Two viewing perspectives were formulated. The first is the newsreader perspective, comparable to the existing situation of the narrator of the Virtual Storyteller [ST08]. The embodied storyteller in the newsreader perspective gazes directly towards the user. Next is the storyteller perspective, derived from existing storytelling situations. An audience is introduced and the camera is placed besides the audience.

The main nonverbal behaviour of the embodied storyteller remains the same and is the gaze behaviour of the embodied storyteller. Existing literature was used to see how gaze behaviours can be simulated by a gaze model. The gaze movement model developed by Fukuyama et al. [FO02] described three important gaze parameters that play an important role in constructing correct gaze behaviour. The three gaze parameters are the gaze length, the gaze amount and the gaze points while not gazing directly towards the user. The gaze model by Mutlu et al. [MF06]. is specifically designed to simulate the gaze behaviours of an actual storyteller. Mutlu et al. used an existing gaze algorithm created by Cassell et al. [TJ97] that utilizes the distributions of the themes and rhemes of an utterance. The distributions are used to infer whether the agent is looking-at or looking-away from the user. The parameters and the notion of using gaze distributions to determine the gaze points are used in the development of the gaze model for the embodied storyteller.

Besides analyzing existing literature, an actual storyteller was annotated to determine the exact gaze points, with their gaze durations and gaze frequencies for the embodied storyteller. The Dutch TV-show "Elly en de wiebelwagen" was annotated. This led to the description of five distinct gaze points for the embodied storyteller in the storyteller perspective. Namely the audience, the left-page of the book, the right-page of the book, the camera and the story-related gazes. The first annotation phase demonstrated how difficult it can be correctly write down gaze behaviours. All resulted in a gaze behaviour lasting one second. A second annotation phase was introduced to annotate the gaze behaviours in milliseconds, which resulted in a increased gaze durations for the gaze at the camera and the story related gazes. Subsequently a news broadcast with an actual newsreader was also annotated, the results of these annotations serve as an indication for the durations and gaze frequencies used by the embodied storyteller from a newsreader perspective.

The information gathered from the annotations, combined with gaze parameters and the formulation of a gaze distribution, led to the construction of a gaze model for the embodied storyteller. The gaze model is divided into three parts. The first part of the gaze model determines the available gaze points at a certain time in the story. For example the embodied storyteller always starts reading from the left-page of the book when at the beginning of the storytelling and ends with a gaze at the camera. The second part of the model selects a gaze behaviour from the available gaze behaviours, this is done by using the amount of gaze at the different gaze points. The final and third part of the model determines the gaze duration. To keep this flexible a duration is selected that is a value between the mean of the gaze and its standard deviation. The embodied agent performs the gaze behaviours during the telling of the story. The gaze model is flexible enough to create different sequence of gaze behaviours every time the model constructs a gaze sequence for the embodied storyteller.

The gaze model was implemented in a Virtual Human called Armandia within the Elckerlyc framework [WR10]. Elckerlyc provided the ability to recreate a compelling storytelling situation as found in "Elly en de wiebelwagen". The newsreader and storyteller perspectives were created by placing the camera at specific locations in the virtual environment. Although Elckerlyc has some restrictions due to the lack in documentation and the use of the non-verbal and verbal behaviours, it was possible to create the an adequate storytelling situation. The end result is an embodied storyteller to telling a tale from two different perspectives, accompanied by a variety of different gaze behaviours.

To assess the different viewing perspectives and to evaluate the gaze model a user survey was conducted. The survey consisted of two video fragments, each being one of the two viewing perspectives (newsreader or storyteller). The fragments lasted around forty-four seconds each. To prevent any influence of the sequence of viewed perspectives, two versions of the user survey were created, the only difference being the sequence in which the perspectives were seen. The survey asked questions regarding the gaze duration and gaze frequency of each of the different gaze points for both viewing perspectives. Also questions were asked regarding the posture and body movements of the embodied storyteller. The user survey ended with the question if the participant had any preference for one of the viewed perspectives.

In total nineteen participants completed the user survey where the storyteller perspective was shown first and fifteen participants completed the user survey where the newsreader perspective was shown first. None of the gaze behaviours or body movements were rated as clearly negative or clearly positive. However the comments of the participants indicated that the gaze behaviours and body movements need some improvements. For example one participant mentioned that the gaze at the right-side of the book looked like a gaze at the floor. As mentioned by Poppe et al. [PR07] is the influence of the current position of the user's viewpoint of the situation. The viewpoint can influences the estimate of the user on correctly stating the agents current gaze. Meaning that the user error might be due to the positioning of the camera for both viewing perspectives. This influence remains to be further investigated. Subsequently another participant remarked that the gaze at the book was too short and that it is impossible to know the entire story by memory.

The objective of this master-thesis was to answer the question if either the newsreader or storyteller perspective is preferred over the other. The user survey shows no clear preference. A small inclination towards the storyteller perspective is seen, but it is not significant. When comparing the results of both surveys, it is interesting to see that the storytelling perspective when seen last, has a clear preference. In contrast to the survey were the storytelling perspective is seen first, which has an even distribution for both viewing perspective with no clear preference. It is clear that there is a huge diversity in people's opinions about correct gaze behaviours and correct storytelling with perfect posture and body movements.

The end result is a first step in analyzing the influence of different viewing perspectives on the viewing experience of the user. The storyteller perspective enables us to add an extra gaze point, in the form of an audience. The participants acknowledged the existence of the audience, even if the audience is not visible to the user. One participant stated that they had the idea that the story was told to a group of children and this made the experience more engrossing. However the storytelling perspective is not clearly preferred over the newsreader perspective. Why is there is no preference? Is the story itself more important than the situation and perspective from which the story is told? Users are influenced by the entire viewing experience, when forming a opinion about specific parts of the experience. Research should not only focus on the behaviours itself, but the viewing experience as a whole.

REFERENCES

[CJ01] J. Cassell, Embodied Conversational Agents Representation and Intelligence in User Interfaces, AI Magazine, Volume 22, No 4, American Association for Artificial Intelligence, p67-83, 2001

[FO02] A., Fukayama, T., Ohno, N., Mukawa, N., Hagita, Messages embedded in gaze of interface agents – impression management with agent's gaze, CHI'02: Proceedings of the SIGCHI conference on Human factors in computing systems, p41-48, 2002, ACM press

[GD02] Givens, D., B., The Nonverbal Dictionary of Gestures Signs and Body Language Cues, Spokane, Washington: Center for Nonverbal Studies Press, 2002

[HD05] Heylen, D., Head Gestures, Gaze and the Principles of Conversational Structure, International Journal of Humanoid Robotics, Volume 3, No. 3, p1-27, 2005

[HK08] Heylen, D., Kopp, S., Marsella, S. C., Pelachaud, C., Vilhjálmsson, H., The Next Step towards a Function Markup Language, In Proceedings of IVA 2008, Volume 5208/2008, p270-280, 2008

[KK06] Kopp, S., Krenn, B., Marsella S., Marshall, A. N., Pelachaud, C., Pirker, H., Thórisson K. R., Vilhjálmsson, H., Towards a Common Framework for Mulitmodal Generation: The Behaviour Markup Language, In Proceedings of the 6th International Conference on Intelligent Virtual Agents, Volume 4133/2006, p205-217, 2006

[LG08] Louwerse, M.M., Graesser, A.C., McNamara, D.S., Lu, S. Embodied Conversational Agents as Conversational Partners, Applied cognitive Psychology, 2008, Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/acp.1527

[MF06] Mutlu, B., Forlizzi, J., Hodgins, J., A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior, In Proceeding of HUMANOIDS'06, p518-523, 2006

[MR11] Mendez Solano, G., Reidsma D., A BML Based Embodied Conversational Agent for a Personality Detection Program, Lecture Notes in Computer Science, 2011, Volume 6895/2011, p468-469, 2011

[PR07] Poppe, R.W., Rienks, R.J., Heylen, D.K.J., Accuracy of head orientation perception in triadic situations: Experiment in a virtual environment, Perception, 36(7), p971–979

[RV02] Ryokai, K., Vaucelle, C., and Cassell, J., Literacy Learning by Storytelling with a Virtual Peer, In Proceedings of Computer Support for Collaborative Learning, p352-360, 2002

[ST08] Swartjes, I., Theune, M., The Virtual Storyteller: Story Generation by Simulation, In proceedings of the Belgian Netherlands Artificial Intelligence Conference (BNAIC) 2008, Boekelo, the Netherlands, 2008

[SV01] Silva, A., Vala, M., Paiva, A., Papous: The Virtual Storyteller, In Intelligent Virtual Agents, 3rd International Workshop on Intelligent Virtual Agents, p171-180, 2001

[TJ97] Torres, E., O., Cassell, C., Prevost, S., Modeling Gaze Behavior as a Function of Discourse Structure, In Proceedings of the First International Workshop on Human-Computer Conversations, 1997

[TK02] Thorisson, K.R., NATURAL TURN-TAKING NEEDS NO MANUAL: COMPUTATIONAL THEORY AND MODEL, FROM PERCEPTION TO ACTION, In Multimodality in Language and Speech Systems, Granström B., House D., Karlsson I., (Eds.). Kluwer Academic Publishers, pp173-207, 2002

[VP08] Vinciarelli, A., Pantic, M., Bourlard, H., Pentland, A., Social Signals, their Function, and Automatic Analysis: A Survey, ICMI'08, ACM, October 20-22, 2008

[WR10] Welbergen van, H. and Reidsma, D. and Zwiers, J. A Demonstration of Continuous Interaction with Elckerlyc, In: Third Workshop on Multimodal Output Generation, MOG 2010, 6 July 2010, Dublin, Ireland.

APPENDICES

Contents of the CD

All of the material used and created for this master thesis is provided on CD. The CD includes the annotations, the software and the results of the user survey. The literature used as background material will also be provided on the CD, see also references section for a full list of used literature.

Annotations

- The first annotation iteration: Ten video fragment were annotated, the data contains the annotations and the average results derived from the annotations for the different gaze behaviours.
- The second annotation iteration: Three of the ten video fragments were annotated using ELAN. The data contains the annotations in ELAN as well as the ELAN files. Subsequently the average results are available.
- News broadcast annotations: Contains the results of two annotations of a newsreader, including the two news broadcasts themselves.

Elckerlyc

- The project with the embodied storyteller, containing the source files, the resources files, libraries and documentation.

User Survey

- Two video fragments used in the user survey, containing one of the two viewing perspectives.
- Both the surveys, with the only difference being the sequence in which the video fragments were seen.
- The average results of the user surveys.
- Comments participants made regarding the gaze behaviours, posture and body movements and the preference for one of the two viewing perspectives.