Feedback Gesture Generation for Embodied Conversational Agents

Malin Bergensträhle

TRITA-NA-E03133
Feedback Gesture Generation
for Embodied Conversational Agents

Malin Bergenstråhle

TRITA-NA-E03133

Master’s Thesis in Computer Science (20 credits)
at the School of Engineering Physics,
Royal Institute of Technology year 2003
Supervisor at Nada was Hercules Dalianis
Examiner was Yngve Sundblad
Abstract

This Master’s Project is carried out in the context of the NECA project. The NECA project develops a system that generates multi-modal dialogues for embodied conversational agents. The aim of this project is to develop and implement an algorithm that generates feedback gestures for the characters in the dialogue. Subsequently the effects feedback gestures might have on the general impression of the dialogue are evaluated. Feedback gestures are performed by a person involved in a conversation but not having the speaking turn at the moment. This person supplies the speaker with feedback so that the speaker knows if s/he has understood or not. The effects were measured in terms of entertainment, naturalness, liveliness and ability to remember the dialogue content.

The algorithm was developed and implemented and an evaluation was carried out. The evaluation results show no big differences in the subjects’ apprehension of the dialogue with feedback gestures compared to the dialogue without feedback gestures. The subjects judged the dialogue without feedback gestures as slightly more entertaining and natural but the one with feedback gestures was judged as livelier.

The memory test shows that the subjects remembered the content of the dialogue without feedback gestures better than the content of the dialogue with feedback gestures.

Generering av feedbackgester för konverserande agenter

Sammanfattning

Detta examensarbete utfördes som en del av NECA-projektet. NECA-projektet utvecklar multimodala dialoger för konverserande agenter. Syftet med detta examensarbete var att utveckla och implementera en algoritm som genererar feedbackgester åt de karaktärer som agerar i dialogen. Därefter skulle också de eventuella effekter som feedbackgesternas hade på dialogen vad gäller underhållningsvärde, naturlighet, livlighet och förmåga att minnas dialogens innehåll fastställas.

Algoritmen utvecklades och implementerades och därefter gjordes en utvärdering av det förändrade systemet. Utvärderingen visar inga stora skillnader beträffande hur försökspersonerna uppfattade dialogen utan feedbackgester jämfört med den med feedbackgester. Utefter försökspersonernas bedömningar var dialogen utan feedbackgester något mer underhållande och naturlig medan dialogen med feedbackgester var livligare.

Minnestestet visar att försökspersonerna kom ihåg innehållet i dialogen utan feedbackgester något bättre än innehållet i dialogen med feedbackgester.
Contents

1 Background .............................. 1
  1.1 The NECA System ......................... 2
  1.1.1 Generation of Scripted Dialogues ........ 2
  1.1.2 The NECA Architecture ................. 3
  1.1.3 Multi-modal Natural Language Generation .. 3
  1.1.4 The M-NLG Module in NECA .......... 4
  1.2 Starting point - Gesture Generation in NECA .. 6
    1.2.1 Choice of Task ...................... 7
    1.2.2 Requirements ....................... 8

2 Theory .................................. 9
  2.1 Descriptive ......................... 9
    2.1.1 Why Gestures? ..................... 9
    2.1.2 Common Taxonomies for Gestures .... 10
    2.1.3 Use of Taxonomies in Gesture Generation 11
  2.2 Computational ...................... 12
  2.3 Empirical ............................. 12

3 Methodology ............................ 15
  3.1 Literature Review ..................... 15
  3.2 Program Design ....................... 15
    3.2.1 PDA's .............................. 17
    3.2.2 Feedback Decider Module .......... 18
    3.2.3 Feedback Gesticone ................. 18
    3.2.4 Temporal Ordering .................. 19
  3.3 Implementation ....................... 20
  3.4 Evaluation Method ..................... 20

4 Results ................................ 23
  4.1 Result of Implementation ............. 23
    4.1.1 Matching Feedback Gestures - feedbackGesticone 24
    4.1.2 Parallel Dialogue Act Generator - ParallelGen . 24
    4.1.3 Feedback Decider - FeedbackDecider ........ 25
  4.2 Evaluation Results .................... 26
4.2.1 How Entertaining were the Dialogues? .................................. 28
4.2.2 How Natural was the Characters’ Behaviour? .......................... 28
4.2.3 How Lively were the Dialogues? ........................................... 29
4.2.4 Results from Memory Testing ............................................. 29
4.2.5 Results from Open-ended Questions .................................... 30

5 Conclusions ................................................................................. 31
  5.1 Entertainment ................................................................. 31
  5.2 Naturalness ................................................................. 31
  5.3 Liveliness ................................................................. 32
  5.4 Memory ................................................................. 33

6 Discussion and Future work ......................................................... 34
  6.1 Discussion on Chosen Task and Method ................................. 34
  6.2 The Relevance of the Conclusions and Future Work ............... 35

7 References ............................................................................... 36

A Gesture generation in NECA ....................................................... 39

B Feedback gesture generation in NECA ........................................ 41

C Flow Charts ........................................................................... 43

D Questionnaire ......................................................................... 45

E Software documentation ........................................................... 52
  E.1 Added Modules .............................................................. 52
  E.2 Changed Modules .......................................................... 55

F Algorithm for calculating probabilities ......................................... 56

G Answers to open-ended questions ............................................. 58
Chapter 1

Background

The underlying idea of Embodied Conversational Agents (ECAs) is that if we can make our computers understand our way of communicating and program them to communicate in the same way as we do, it will make the interaction between us and the computers more intuitive. In a distant future we might consider computers more like intelligent friends rather than just tools.

GESTURES are an important part of the natural behaviour of humans conducting a dialogue. A dialogue can be seen as a flow of information between two humans. When we humans want to give or require information we express ourselves in multiple modalities. Gesturing is one such modality. To construct ECAs that behave in a human way it is therefore necessary to make them gesture. This gesturing should not be random, but serve the same purposes that human gesturing does, to complement speech when communicating.

This Master’s project is carried out in the context of the NECA\(^1\) project that develops a system generating Scripted Dialogues for animated agents. A Scripted Dialogue is here defined as “a dialogue performed by two or more embodied agents on the basis of the description of that dialogue” (Fiwek and van Deemter, 2002). The Scripted Dialogue is generated from a less detailed, abstract description of the dialogue, a \emph{Scene Description}. One advantage with generating Scripted Dialogue contra real time generated dialogues is the possibility to manipulate the dialogue with techniques known from writing manuscripts for theatre plays. One such technique is to illuminate some important information through letting the characters repeat it between each other. In the prototype scenario of this project Scripted Dialogues are generated in the car sales domain. The user is provided information about some car through watching a dialogue between a car seller and a car buyer. The aim of this dialogue is to keep the users’ interest for the dialogue alive throughout the conversation and to mediate some specific information about the car, that the user has specified that s/he is interested in (e.g. environmental issues, horse power etc.). Since the characters performing the dialogue use multiple modalities, a part

\(^{1}\)NECA stands for “Net environment for Embodied emotional Conversational Agents”. For further information see http://www.ai.univie.ac.at/NECA/project/project.html
of the generation of the Scripted Dialogue is gesture generation.

The aim of this thesis is to propose and implement an algorithm for *feedback gesture generation* for the embodied conversational agents in NECA and to investigate the effects of feedback gestures on the impression of the dialogue according to entertainment, naturalness of characters’ behaviour and liveliness. A feedback gesture is performed by a person involved in a conversation but not having the speaking turn at the moment. This person supply the speaker with feedback so that the speaker knows if s/he has understood or not. The purpose with adding feedback gestures to the dialogue is to make the dialogue look more natural and entertaining. The problem definition can be formulated as

*Given the content of a Scene Description, how can we realize feedback gestures and which influence does the presence of feedback gestures have on the total impression of the dialogue according to entertainment, naturalness of characters’ behaviour and liveliness?*

The section 1.1 give a description of the NECA system and in section 1.2 the gesture generation problems in NECA is formulated and some requirements for the extension of the system are stated. In the following chapter 2 the theoretical ground for gesture generation is briefly discussed. Chapter 3 consists of a methodology description of this project. The results are presented in chapter 4 and some conclusions are drawn in chapter 5. The project is discussed from a personal point of view in chapter 6 and there are also some suggestions for future work given.

## 1.1 The NECA System

The NECA project focuses on generation of Scripted Dialogues that are to be performed by two or more embodied agents. The generation includes both natural language generation (NLG), generation of nonverbal cues such as gestures as well as visual and audial realization of the dialogue. The generation of Scripted Dialogue is introduced in 1.1.1. In the following section 1.1.2 the architecture of the NECA system is presented and in section 1.1.3 a general introduction to Multi-modal Natural Language Generation is given. Finally, in section 1.1.4, the Multi-modal Natural Language Generator in NECA, the M-NLG module (Piwek, 2003), is described.

### 1.1.1 Generation of Scripted Dialogues

Here, a Scripted Dialogue is a dialogue performed by two or more embodied agents on the basis of the description of that dialogue. The Scripted Dialogue generation includes generation of the *Dialogue Script* and the execution of it. The Dialogue Script contains all information about the actions needed for the agents to be able to perform it. The NECA system generates Dialogue Scripts for two prototype scenarios. *Socialite*, is a multi-user application for the social domain where users create their personal agents and then let them live in a virtual environment. The agents report back to the user about encounters with other agents and so on. The
second application, *eShowroom*, is the one in focus of this thesis. It generates a
dialogue between a buyer and a seller in the car sales domain. The aim of generating
this kind of dialogue is to develop a prototype of a new showroom for the products
of online shops. The visitors of the showroom will receive product information by
watching the dialogue between the embodied agents. Important information can be
highlighted through for example repetition of that information in the dialogue. The
user’s influence is limited to setting some parameters before the Dialogue Script is
generated. The parameters describe what is important about the car for the buyer
character (e.g. sportiness, environmental issues) and also the personality traits of
the agents (e.g. politeness, extroversion).

1.1.2 The NECA Architecture

There is a flow of information in the NECA system. The flow of information begins
with that the user enters input in form of specifying the personality traits and defining
the customer’s interests according to the car. Through these input values the
characters are set and some dialogue goals are specified. Additionally, the generator
takes as input a database with facts about the chosen car (price, speed, horse power
etc.) and a database that correlates facts with value judgements. The information
flows subsequently through a pipeline process shown in Figure 1.1 where each mod-
ule adds information and the result is a detailed final description of the dialogue.
This description is player independent. The final step is to convert it to a control
sequence that can be understood by some media player. In the eShowroom proto-
type we use Microsoft Agents\(^2\) (MSAgents) as media player whereas Macromedia
Flash\(^3\) is used for Socialite.

1.1.3 Multi-modal Natural Language Generation

Multi-modal Natural Language Generation (M-NLG) is the process of constructing
a linguistic output for multiple modalities from a non-linguistic input. The
non-linguistic input is often an abstract description of some content. The M-NLG
generator has to transform this description to speech and gestures that represent the
same content as the input. The generation has to go through some steps. The verbal
part is often generated in the following way (Jurafsky and Martin, 2000). First, a
lexical selection is made to choose the lexical item most appropriate for expressing
particular concepts. Next step is to build the sentence structure through combining
the verbal lexical items into linguistic sentences. Finally the discourse structure is
built up by putting the sentences together in a coherent way. Note that gestures also
can be lexical items. Sometimes they are lexical items themselves, like a nod means
yes or head shake means no. In other cases they must appear in combination with
some verbal lexical item to get the correct conceptual meaning. The representation
of the lexical items expressed with other modalities can be separately generated or

\(^2\)http://www.msagent.org/
\(^3\)http://www.macromedia.com/
generated together with the verbal structure. The final output is a suggestion of how to realize the given input information by using verbal natural language and gestures.

1.1.4 The M-NLG Module in NECA

The gesture generation takes place in the Multi-modal Natural Language Generation (M-NLG) module in Figure 1.1. The input to this module is a given content of a so called Scene Description. A Scene Description is an abstract description of the interaction, containing information about the subject matter of the interaction, the personalities of the characters, their common ground \(^4\) etc. The personalities of the characters are regulated by numerical values of \textit{politeness, agreeableness} and \textit{extraversion}. These values are defined in the interval \([0, 1]\), where for example \textit{politeness} = 1 means a very polite person whereas \textit{politeness} = 0 means an impolite person. The Scene Description consists also amongst others of a set of Dialogue Acts (DA:s). The Dialogue Acts are “units” with information about Dialogue Acts. Typical values of the Dialogue Act type attribute are \textit{inform, question} or \textit{feedback}. The

\(^4\)Two persons' common ground is the sum of their mutual, common or joint knowledge, beliefs and suppositions, see Clark, 1996
other attributes are speaker, addressees, semantic content and other specifications, see Figure 1.2 for a pictorial description of a Dialogue Act and a Person. The notion in the rectangles in this figure are attributes connected to Person and Dialogue Act. In the bubbles connected to the attributes, names of some example values of the attributes are given. The Dialogue Acts’ temporal order is separately represented.

![Diagram of Person and Dialogue Act with their attributes]

*Figure 1.2. The Foundation stones Person and Dialogue Act and their attributes*

in the Temporal Ordering element and it is possible to change it in a later phase. The Temporal Ordering element is just a container of Dialogue Act identification variables in the temporal order they are going to be performed. The order of the Dialogue Acts is then allowed to be changed without explicitly changing the content of the Dialogue Acts themselves. The task for the M-NLG module is to use information in the original Dialogue Acts in the Scene Description to construct the verbal dialogue and assign proper gestures for the characters to perform. The generated information is then added to the Dialogue Acts and the merged information is passed to the next module.

The Scene Description and the Dialogue Script are represented in the XML-compliant RRL\(^5\) (Rich Representation Language). The information is represented in RRL throughout the whole NECA pipeline process until it reaches the media player. The difference between the Scene Description and the Dialogue Script is obvious when looking at the RRL-representations of them. In the Scene Description we find among others a specification of which the dialogue participants are, the temporal order of the Dialogue Acts and a set of Dialogue Acts. The information within the Dialogue Acts is quite poor and abstract before the M-NLG generation. Basically the only thing that is specified is who is going to say what in terms of Dialogue Act type and semantic content. In Figure 1.1.4 there is an example of the RRL representation of a DA in a Scene Description. After completed Multi-modal Natural Language generation, the big difference of the Dialogue Script compared to the Scene Description is the appearance of the Dialogue Acts. There might be some added Dialogue Acts and their internal temporal order might have been changed in

\(^5\)A framework for information representation in NECA, see http://www.ai.univie.ac.at/NECA/RRL
<dialogueAct id="DA1">
  <domainSpecificAttr type="positiveEvaluation"/>
  <speaker id="ritchie"/>
  <addressee id="tina"/>
  <semanticContent id="d_3">
    <unaryCond arg="x_1" id="c_3" pred="safe"/>
  </semanticContent>
</dialogueAct>

**Figure 1.3.** Example of DA specification in Scene Description

the Temporal Ordering element but the main difference is that now, the Dialogue Acts contain a multi-modal realization of the earlier abstract content description of the dialogue. The same Dialogue Act as shown in Figure 1.1.4 can have the appearance as in Figure 1.4. Basically this corresponds to that Ritchie is going to

<dialogueAct id="DA1">
  <domainSpecificAttr type="positiveEvaluation"/>
  <speaker id="ritchie"/>
  <addressee id="tina"/>
  <semanticContent id="d_3">
    <unaryCond arg="x_1" id="c_3" pred="safe"/>
  </semanticContent>
  <gesture id="G5" identifier="Blink1"/>
  <sentence id="s_mnlg_19">It is terribly safe.</sentence>
  <gesture id="G6" identifier="LookAddressee"/>
</dialogueAct>

**Figure 1.4.** Example of DA specification in Dialogue Script

say “It is terribly safe’ to Tina while blinking and looking at her.

## 1.2 Starting point - Gesture Generation in NECA

MSAgents is the mediaplayer used in eShowroom and it has its limitations concerning gestures. The generation should disregard this fact and generate a player independent script possibly more detailed than the MSAgents are able to realize. Still, a limited set of gestures is the basis of the gesture generation. The generator is mapping a suitable gesture to each Dialogue Act type. Basically it is based on question - inform gestures. When a character is asking a question s/he always rises her/his hand into the air and when giving some information the character throws out one arm a bit lower than when asking a question. There are some variations in their
gazing behavior. For example, the characters look at each other or at the car and sometimes they are frowning. Some extraordinary movements to make the dialogue more interesting are also included, for example Ritchie sometimes laughs with the hand in front of his mouth or Tina takes her glasses on. A complete description of all movements and which DA types they are connected to can be found in Appendix A. The agents are speaking one at a time and while one of them is speaking the other agent is waiting for his/her turn without doing any gestures. A typical sequence in a NECA dialogue could look like in Figure 1.5. This is what happens:

1. Tina asks: “How much does it consume” while raising her hand into the air. Ritchie does nothing at all while Tina is speaking.

2. Ritchie says: “It gets 34 miles per gallon” while throwing out his right arm. Tina waits for Ritchie to finish speaking.

![Figure 1.5. A sequence of DAs](image)

1.2.1 Choice of Task

To delimit the topic of this thesis, we are focusing on adding generation of so called feedback gestures to the NECA system. As mentioned, a feedback gesture is performed by a person involved in a conversation but not having the speaking turn at the moment. This person supplies the speaker with feedback so that the speaker knows if s/he has understood or not. Feedback can also be an emotional reaction on what the speaker just said or an expression of some opinion about the content of the utterance. Feedback is crucial for the success or failure of a spoken language dialogue (Allwood et al., 1993). The feedback is either spoken (“yes”, “no”, “uh-huh”, “what”), gesticulated (head nod, head shake, shoulder shrug) or a combination of both. A feedback expression serves one of the following purposes (Cerrato, 2002):

**Show continuation** the interlocutor has understood the message and wishes to show continuation of contact. S/he expects more information from the speaker.

Possible feedback signs are head nod, smile, short expressions (“yeah”, “uh-huh”).
**Acknowledge comprehension** the interlocutor has understood the message and wishes to let the speaker know that. The reason for this is either that the speaker requires feedback through some signal or just that the interlocutor her/himself wishes to announce that. Possible signs are head nods, gazing behaviour, smiles or repetition of a part of the speakers utterance.

**Express a point of view** the interlocutor agree/disagree with the utterance. S/he wishes to announce her/his point of view and does this for example by shake her/his head, nod, make a facial expression or say something briefly.

To make a dialogue human resemble feedback behaviour is probably necessary, and that is what this project will investigate. The contribution of this thesis to NECA is

> Given the content of a Scene Description, how can we realize feedback gestures and which influence does the presence of feedback gestures have on the total impression of the dialogue according to entertainment, naturalness of characters’ behaviour and liveliness?

### 1.2.2 Requirements

The first thing was to state some requirements for the extension of the system that was going to be made. These requirements on the modified gesture generation model are here stated and explained.

**Easy to extend** It is desirable that adding new gestures to the model can be carried out through reading a software documentation and that the programming style is clear and understandable by others.

**Timing** The feedback cues should be well synchronized.

**Maximize use of available information** Since there is a lot of information in the Dialogue Acts, specify the choice of gestures with respect to this information to improve variation in performance. We are creating a Scripted Dialogue and could also use global dialogue information so that every given cue, verbal or nonverbal, is well motivated to be performed.

**Use empirical results** To make behaviour look natural, use statistical data from transcriptions of human-human dialogues to assign probabilities to gestures.
Chapter 2

Theory

Literature on topics useful when generating feedback gestures for Embodied Conversational Agents is various and belongs to several different disciplines. We need some knowledge about gestures from a psychological and linguistic point of view and we also need to investigate the techniques of implementing gestures into ECAs.

In the following chapter there are three different kinds of literature described. This is a summary of the material in a more extensive literature review and it sums up some important results from earlier work. The three different kinds of literature are Descriptive, discussed in section 2.1, Computational, considered in section 2.2 and Empirical in section 2.3.

2.1 Descriptive

The descriptive literature tries to explain how, when and why humans are gesticulating. In order to explain different kinds of gestures and their meaning there are different ways of categorizing gestures. We are interested in feedback gestures, i.e. gestures that are performed as reactions on an utterance. Do these gestures fall into the same category in some taxonomy or are they members of different categories?

2.1.1 Why Gestures?

As mentioned, gestures are an important part of the natural behaviour of humans conducting a dialogue. It is commonly assumed that gestures and speech both origin from the same “idea unit” in human thought and then somehow they bifurcate to different motor systems (McNeill, 2000). Either we use speech only or gestures only as information channel or we gesticulate and speak simultaneously. In the latter case we can either do gestures that mean the same thing as the speech (e.g. nodding while saying “yes”) or the gestures can add some information to the speech (e.g. simulate a spiral with your fingers while saying “she went up the stairs”, the gesture adds the information “it was a winding staircase”). After a dialogue, the speaker and listener are not always aware of the gestures that turned up during the conversation. In fact, most speakers are not aware that they have been using their hands at all. Neither
the listeners have a memory of the gestures but they have perceived them and the information that they gave (Cassell, 2000). People gesture even if the listener is not present, e.g. when speaking on the telephone (Rimé, 1982). This fact shows that gestures are tightly bound to human speech. Our natural way of communicating is multi-modal. Although we can also understand information provided by only one modality, multiple modalities enriches the communication and we can express ourselves in a refined way.

To sum up, there is a complex connection between speech and gestures. To make ECAs act human resemble we have to model this connection. We want the agents to make gestures that look natural when they speak so that the combined information from the different modalities makes sense to us. Additionally, the agent/s not speaking at the moment should be able to react not only on what is said but also on nonverbal cues. For instance, if the speaker looks at a listener to require a feedback signal the listener should nod.

2.1.2 Common Taxonomies for Gestures

Gestures do not occur randomly during a conversation, there is a meaning attached to them. A gesture can be considered as a kind of signal (Clark, 1996). Signaling can be performed in different ways but basically there are three different kinds of signals. These kinds are:

**Icons** are demonstrative gestures that depict what is being talked about or a common metaphor for the entity. These gestures are tightly bound to the speech, for example depicting the shape of a birthday cake while talking about it. Iconic gestures tend to anticipate the words they go with, the gesture comes shortly before the entity or action is described with words (Clark, 1996).

**Indices** are gestures physically connected to the entity the refer to, e.g. pointing gestures.

**Symbols** are gestures with their own symbolic meaning; emblems, e.g. head nod or thumb up.

When we look at feedback gestures from this point of view, we see that they principally can be of any of these kinds but the *symbol* signs seem to be the most common one.

One opinion about gestures is that it is not possible to put gestures into distinct categories (McNeill, 2000). This does not violate the idea that gestures are signals and some fundamentally different kinds of gestures are sorted out in a taxonomy called *Kendon’s continuum* (Figure 2.1).

![Figure 2.1. Kendon’s Continuum](image-url)
Gesticulation is spontaneous movements that accompanies speech, e.g. hand rising upward while the speaker says and “he climbs up the pipe”

Language-Like gestures are similar to gesticulation but they replace some word in an utterance. They do not alone have a specific meaning but in some context they can replace words. “it takes ten minutes to get there” [while making a pair of walking legs with two fingers]. The fingers replace the words “by foot”.

Pantomimes are gestures that depict objects or actions, but speech is not obligatory. They might be accompanied with some sound effect “oopp” or “click”.

Emblems is the same concept as what earlier was called symbols. They are movements with a special meaning within some culture or community, for example thumb up, head nod, V-sign etcetera.

Signed Languages are full-fledged linguistic systems, speech is not needed at all.

Walking from left to right in the continuum we see that obligatory presence of speech declines, that the gestures receive more language like properties and that idiosyncratic gestures are replaced by socially regulated signs. Gestures involved in common human conversations belong mainly to the leftmost part of the continuum, that is gesticulation (McNeill, 2000).

The approach to divide gesticulation into different categories (e.g. icons, indices and symbols like above) might not be fully applicable (McNeill, 2000). It is proposed that the gestures should instead again be seen as a continuum with more or less “iconicity”, “indicity” or “symbolicity”.

2.1.3 Use of Taxonomies in Gesture Generation

There seem to be some fundamental features in gesture classification. One is that symbols (emblems) probably is a totally different kind of gesture compared to spontaneous gesturing, although they occur often during conversations. In a generation model it would be preferable to generate emblems together with the language. The reason for this is because the emblems have very specific meanings and the function of them are the same as words or whole sentences. An emblem can be performed instead of a sentence while other kinds of gesticulation must be accompanied by speech, signed languages of course excluded.

Another fundamental feature is that the main part of the conversational gestures seems to be spontaneous and often unwitting (Cassell et al., 2000). While dealing with the spontaneous gestures one must choose a taxonomy and specify some properties of the context in which the gestures appear. The task is then to match the generated text with these properties and to decide how often a specific character is gesticulating. The focus of this thesis, feedback, is a conversational phenomena rather than a gesture category and this phenomena can include gestures from different categories in the taxonomies mentioned earlier. Feedback gestures are not likely to be accompanied by speech, since that would interrupt the person talking.
If they are accompanied by some utterance ("yeah", "yes", "uh-huh" etc.) the movements are probably redundant and consists of an emblem with the same meaning as the utterance but if not the feedback gestures are more like gestulation.

Feedback is content related because it is a reaction on something that is said. Icons and index gestures or gestures with a lot of "iconicity" or "indicity" might not be commonly used as feedback gestures, whereas symbols or emblems like head nods and head shake are more frequently used as feedback.

2.2 Computational

The literature studied here was mostly descriptions of earlier implementations of Embodied Conversational Agents and for all their gestures. Implementations that were studied were URBAN (Nordstrand, 2001), BEAT (Bickmore et al., 2001), REA (Cassell, 2001) and AutoTutor (Rajan et al., 2002). They all had slightly different approaches to this subject. One approach is to study the actual movements of hands and arms and then give hands and arms exact positions in space at the same time as speech is generated. This approach is not applicable on this thesis since we have no possibilities to choose exact positions of body parts. In deed, we could in the Dialogue Script include such kinds of specifications. The translation of the Dialogue Script to a control sequence understandable by the media player, MSAgents, would then have to disregard these specifications since the gestures are represented in MSAgents as a limited set. The MSAgents used in the current implementation have a set of about 30 gestures. The other approach is to choose some suitable movement or combination of movements from a collection and this is exactly the prerequisites of this project. Also within this approach different solutions can be found. The approach finally chosen for the implementation has been inspired by an approach we call The functional approach. The functional approach is basically the idea behind REA (Cassell et al., 2001), an agent who interacts with the user and whose domain of expertise is real estate in Boston.

Here, the main goal is to identify the fundamental phases and high level structural elements that make up a conversation. Elements are described in terms of their role or function in the exchange and Table 2.1 (Cassell et al., 2001), shows that the same behaviour can fill several functions.

2.3 Empirical

This section describes some results from empirical studies on human gestures that are useful when implementing gesture generating algorithms for conversational agents. The agents are supposed to act like humans and therefore we first need to know how humans actually do when they gesture. One way of getting that knowledge is to study real humans conducting dialogues. We need to discover which variables
<table>
<thead>
<tr>
<th>Communicative Functions</th>
<th>Communicative behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initiation and termination</strong></td>
<td></td>
</tr>
<tr>
<td>React to new person</td>
<td>Short glance at other</td>
</tr>
<tr>
<td>Break away from conversation</td>
<td>Glance around</td>
</tr>
<tr>
<td>Farewell</td>
<td>Look at other, head nod, wave</td>
</tr>
<tr>
<td><strong>Turn-Taking</strong></td>
<td></td>
</tr>
<tr>
<td>Give Turn</td>
<td>Look, raise eyebrows (followed by silence)</td>
</tr>
<tr>
<td>Want Turn</td>
<td>Raise hands into gesture space</td>
</tr>
<tr>
<td>Take Turn</td>
<td>Glance away, start talking</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
</tr>
<tr>
<td>Request feedback</td>
<td>Look at other, raise eyebrows</td>
</tr>
<tr>
<td>Give feedback</td>
<td>Look at other, nod head</td>
</tr>
</tbody>
</table>

**Table 2.1. Interactional functions vs behaviours**

that decide when and how a person is gesturing. Is the current role in the conversation important? Does knowledge about the conversation topic impact the gestures? Many questions needed to be answered.

How are empirical studies on gesturing made? The most common method seems to be videorecording a dialogue and then transcribe the gestures and the speech separately, mostly for some specific features. Different methods for transcription are used.

Some of the studies on human gesturing contained information that could be used in some way when designing the gesture generation algorithm. Some transcriptions from the Gothenburg University Spoken Language Corpus (Allwood, 1999) has been investigated with focus on feedback (Cerrato, 2002). The dialogues considered are “information seeking” exchanges. These are similar to the scenario in eShowroom since they are dialogues between a “customer”, asking for information, and a “seller” providing information. Cerrato presents the results about the distribution of feedback. One important result is that customers are much more likely to give feedback than sellers. More than 50% of the customers’ turns consisted of feedback expressions whereas only around 20% of the sellers’ turns included feedback.

The study done for the AutoTutor project, involved students participating in tutoring sessions with tutors (Rajan et al., 2002). The videotaped segments were evaluated in terms of the locations within the tutor’s and the student’s dialogue moves, with focus on feedback. There were four kinds of back-channeling feedback detected: positive, neutral, negative and head nods shown in the Table 2.2. The conclusion was that both students and tutors largely avoided negative and neutral feedback. The students’ feedback were mostly head nods whereas the tutors’ mainly gave spoken feedback.

Some of the statistical results from these investigations were finally used in the implementation. How the results were adapted is described later in section 3.2.2.
<table>
<thead>
<tr>
<th>Kind of feedback</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>yeah, yes, OK, right</td>
</tr>
<tr>
<td>neutral</td>
<td>uh-huh</td>
</tr>
<tr>
<td>negative</td>
<td>no</td>
</tr>
<tr>
<td>head nods</td>
<td>up-down, down-up</td>
</tr>
</tbody>
</table>

Table 2.2. Examples of feedback behaviours
Chapter 3
Methodology

The limited task was to generate some feedback gestures in the context of the generation of NECA dialogues. At the starting point the agent listening at the moment was set to wait for the speaking agent to finish his/her speaking turn. The problem could now be divided into subproblems. Step one was to make a literature review, described in section 3.1 and summarized in chapter 2. Afterward, the existing flow of information in the NECA system was studied in detail and a flow chart for the modules to be added was designed. This step is described in section 3.2. Next, the modules were integrated into the system with the method discussed in section 3.3. The purpose with adding feedback gestures to the dialogue was to make the dialogue look more natural and entertaining. To investigate if the modified dialogue fulfilled these expectations, the dialogue was evaluated by two groups of people. A description of the evaluation method is presented in section 3.4.

3.1 Literature Review

An extensive literature review, summarized in chapter 2, was the initial step. Literature on the psycho linguistic view of gesture was studied as well as papers on computer scientists different aspects on how to deal with the implementation of gestures for embodied agents. It was also important to study deliverables from the NECA project to get to know the system and the ideas behind the project.

3.2 Program Design

After the literature review, the knowledge basis about implementing gestures for embodied conversational agents was stable enough to form an idea about how to implement the feedback gesture generation in the NECA system. Through studying the information flow in the existing Multi-modal Natural Language Generator, the first step in the program design was to add the modules of feedback gesture generation to this flow of information.
One of the basic ideas behind NECA is the idea of Scripted Dialogues. The point is that these abstract descriptions of the dialogues contain information independent of which specific media player that is going to play the dialogue to an audience. Important is that they contain enough information to play a dialogue. The premier task was therefore to generate an abstract description of the feedback behaviour and then let the system translate the description to the media player placed at disposal, the MSAgents. The input description of the dialogue is represented as a Scene Description (in terms of a RRL document, see section 1.1.3) that contains a set of Dialogue Acts (DA:s) and their internal temporal order as a separate element. Each Dialogue Act is associated with a DA type, a speaker and an addressee. All kind of action specification (speech and gestures) that is added to this DA is supposed to be performed by the speaker. Instead of changing the information representation inside the DA:s by specifying “who is doing what” the idea was to create another separate DA, a Parallel Dialogue Act (PDA). The idea is illustrated in Figure 3.2 and PDA:s
are more extensively discussed in section 3.2.1. Henceforth, when referring to DA it means the Dialogue Act containing the specifications for the speaker and the PDA means the Parallel Dialogue Act parallel to the DA in time but storing information about the listener gestures only. The result of this first design phase was a flow chart. The conclusion was that three new modules needed to be created and one was necessary to change. A simplified version of the flowchart is shown in Figure 3.3 and the detailed version of it can be found in Appendix C.

![Flowchart](image)

**Figure 3.3.** New Flow chart

**To be added**

**Parallel Dialogue Act generator** in which the PDA:s are created.

**Feedback Decider** in which the decision about whether a feedback gesture is to be performed or in a PDA or not is taken.

**Feedback Gesticon** in which the choice of gesture for each PDA is made.

**To be changed**

**Temporal ordering element** needs to be changed so that it places each PDA parallel in time to its corresponding DA.

### 3.2.1 PDA:s

The PDA stores all information about what the listener is doing while the speaker is performing his/her DA. Since the PDA is a reaction on the DA, the generation of the PDA is based on information about the DA. An exemplification of this is that if the seller Ritchie says “this car is terribly safe” the buyer Tina nods and smiles. Within the PDA:s the feedback gestures are specified and the agent performing them is always the listener. In case of multiple listeners the system would be easy to
extend to generate feedback gestures for each listener. Each listener would then be provided a PDA for each DA in the dialogue and their gestures would be separately generated. In Figure 3.2 the structure of a DA and a PDA is illustrated. In the example a DA called $DA1$ in which Ritchie says “it has airbags” is parallel to a PDA, $PDA1$, in which the contemporal specification for the listener Tina is two gestures, $G5$ and $G6$.

### 3.2.2 Feedback Decider Module

This module is modelling the frequency of feedback gestures and adapt empirical results to make the generation as realistic as possible. The first task in the Feedback Decider module is to decide whether a feedback gesture is to be performed or not. This decision is based on two assumptions. One is that some DA types are not likely at all to elicit feedback and the other is that the buyer is more likely to give feedback than the seller. The explicit formulation of the first assumption is that the PDA to a DA of one of the types

- **Feedback**
- **Opening Complaint Response**
- **Refuse Answer Response**

is not likely to contain a feedback gesture. The motivation of this assumption is that these DA types are feedback themselves. The probability that a specific character should give feedback is adapted, depending on how many DAs of these forbidden types there are in the dialogue. The exact formulation of the algorithm that makes this calculation is presented in *Appendix F*.

The motivation of the second assumption, that the buyer is more likely to give feedback than the seller, is that the dialogue in NECA can be described as an “information seeking exchange” where the buyer seeks information about the car and the seller provides it. Statistical evidence for the feedback distribution in this kind of dialogues was found and the results (Cerrato, 2002) were used in the feedback gesture generation in NECA as shown in Table 3.1. The random but statistically based decision whether a PDA should contain a feedback gesture or not is made in the Feedback Decider module. In the case the Feedback Decider gives green light the subsequent task is to decide which gesture to be performed and this is the aim of the Feedback Gesticon module. If Feedback Decider decides that no gesture shall be performed, the listener is set to wait for the speaker to finish his/her DA.

### 3.2.3 Feedback Gesticon

The feedback gesture generator Feedback Gesticon is designed with a structure similar to the functional approach (see section 2.2.1). We recall that a DA has a DA type (e.g. *request Value, agree, negative response*) and that the buyer and the seller both have personality values (e.g. *politeness, agreeableness*). The choice of gesture
<table>
<thead>
<tr>
<th>Feedback gestures in</th>
<th>Buyer</th>
<th>Seller</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerrato, 2002</td>
<td>≥ 50 %</td>
<td>20-40 %</td>
</tr>
<tr>
<td>NECA</td>
<td>60 %</td>
<td>30 %</td>
</tr>
</tbody>
</table>

Table 3.1. Adaptation of empirical results. The probabilities given in Cerrato was "Number of individual turns that contained feedback expressions." The probabilities in NECA is coherent with this result. They represent the probability that the Feedback Decider module decides that a PDA shall contain a feedback gesture.

<table>
<thead>
<tr>
<th>Dialogue Act Type</th>
<th>Personality</th>
<th>Feedback Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greeting</td>
<td>polite</td>
<td>smile, nod</td>
</tr>
<tr>
<td>Greeting</td>
<td>impolite</td>
<td>look sceptical, nod</td>
</tr>
<tr>
<td>Request value</td>
<td>polite</td>
<td>thinking gesture</td>
</tr>
</tbody>
</table>

Table 3.2. Examples of feedback gesture assignments

is mostly based on the combination of DA type and personality values. The combination of these values for the current DA is matched with one or more suitable gestures that are fetched from a collection of feedback gestures. Some examples of these matchings are given in Table 3.2 and the complete matching schema is presented in Appendix B. If there are more than one suitable gesture, the choice between them is random. The gestures are given descriptive names such as nod, smile or fold.

3.2.4 Temporal Ordering

When a Parallel Dialogue Act is created, it must be added to the Temporal Ordering element. This element only contains information about in which order the DAs are to be performed and it is represented in the Scripted Dialogue as in Figure 3.4. Now, we define the parallelism of each DA and its corresponding PDA in the Temporal Ordering element as in Figure 3.5.

```xml
<temporalOrdering>
  <seq>
    <act id="DA1"/>
    <act id="DA2"/>
    <act id="DA3"/>
  </seq>
</temporalOrdering>
```

Figure 3.4. Example of Temporal Ordering element in Scripted Dialogue, without PDA

19
<temporalOrdering>
  <seq>
    <par>
      <act id="DA1"/>
      <act id="PDA1"/>
    </par>
    <par>
      <act id="DA2"/>
      <act id="PDA2"/>
    </par>
    <par>
      <act id="DA3"/>
      <act id="PDA3"/>
    </par>
  </seq>
</temporalOrdering>

Figure 3.5. Example of Temporal Ordering element in Scripted Dialogue with PDA.

The output is a revided Temporal Ordering element with the order or the DA:s and the PDA:s.

3.3 Implementation

After the design phase the modules were programmed, integrated into the NECA system and tested. The system was studied in detail before starting the implementation. The code is written in Prolog. The feature structures of the dialogue acts, the persons etc in NECA are represented in ProFIT, an extension of Prolog. The acronym ProFIT stands for Prolog with Features, Inheritance, and Templates (Erbach, 1995).

3.4 Evaluation Method

Finally an evaluation was designed and performed to find out whether the addition of feedback gestures to the NECA dialogue had the wanted effect on the dialogue or not. The aim of the evaluation was to investigate four hypotheses:

1. Audience find a dialogue with feedback gestures more entertaining than a dialogue without feedback gestures.

2. Humans find the behaviour of characters in a dialogue with feedback gestures more natural than the behaviour of characters in a dialogue without feedback gestures.
3. A dialogue with feedback gestures is interpreted as more lively than a dialogue without feedback gestures.

4. It is easier to remember the content of a dialogue with feedback gestures than the content of one without feedback gestures.

The manipulated variable of the evaluation is the presence of feedback gestures. The dependent, measured variables are:

- *entertainment*
- *naturalness of characters’ behaviour*
- *liveliness of dialogue*
- *ability to remember the content of a dialogue*

The subjects were Swedish students and researchers divided into two groups of six persons in each group. The groups were shown two dialogues, one dialogue with feedback gestures ($DFB+$) and one dialogue without feedback gestures ($DFB-$). The course of events of the two evaluation sessions is shown in Table 3.3. The purpose of the questions after each dialogue was to confirm or reject the hypotheses. The subjects were given questionnaires without formulations of the questions (*Appendix D* contains the questionnaire and the exact formulations of the questions). The dialogues were shown on a screen to all subjects in each group at the same time and after each dialogue the questions were projected on the screen, one at a time. Questions, part 1, consisted of:

1. Statements about the dialogue that the subjects were supposed to agree or disagree to.

2. Questions about some details that were mentioned in the dialogue that the subjects were tested if the remembered.

3. One open-ended question about the gesticulation of the characters.

After answering Questions, part 1 the subjects were shown a second dialogue, differing from the first one only according to the presence of feedback gestures. Finally Questions, part 2 were shown and it consisted of:

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$DFB+$ was shown</td>
<td>$DFB-$ was shown</td>
</tr>
<tr>
<td>2</td>
<td>Questions, part 1</td>
<td>Questions, part 1</td>
</tr>
<tr>
<td>3</td>
<td>$DFB-$ was shown</td>
<td>$DFB+$ was shown</td>
</tr>
<tr>
<td>4</td>
<td>Questions, part 2</td>
<td>Questions, part 2</td>
</tr>
</tbody>
</table>

Table 3.3. Course of events in the evaluation sessions
1. The same statements as in part 1.

2. Another open-ended question about the subjects’ apprehensions of the difference between the two dialogues.
Chapter 4

Results

There are basically two different results to present as outcome of this project. One is the actual code of the modified system and the other is the results from the evaluation of the modified system. The code is briefly discussed and presented in section 4.1 and in section 4.2 the results from the evaluation are described. One obvious difference is the appearance of the dialogue. The same sequence of Dialogue Acts as shown when discussing the starting point looks a bit different. Recall that the listener in the original dialogue was inactive while waiting for the speaker to finish. Now, the sequence of Dialogue Acts look like in Figure 4.1. The difference between this sequence of two DA:s together with their PDA:s is that the listener does some movements. Ritchie makes a thinking gesture as a reaction on the first DA and Tina smiles as a reaction on the second DA.

![Figure 4.1. A sequence of two Dialogue Acts and their Parallel Dialogue Acts](image)

4.1 Result of Implementation

The extended NECA system can be seen as the actual product of this project. In Appendix E there is a software documentation of the code added to the NECA system to make it generate feedback gestures. The ideas presented in the section
about program design (section 3.2) were implemented and the modules as they look like implemented in Prolog are here described.

### 4.1.1 Matching Feedback Gestures - feedbackGesticon

The module feedbackGesticon takes care of the choice of gesture for each PDA. It consists of many cases of combinations of different values of DA type and personality values. The DA types of the preceding and following DA are also sometimes controlled to state the context correctly. The principle is to match a gesture with the current DA type, personality and context. Recall the matching example Table 4.1 and Appendix B for a complete matching description. An example of what a typical predicate in feedbackGesticon looks like is given in the code sample in Figure 4.2.

<table>
<thead>
<tr>
<th>Dialogue Act Type</th>
<th>Personality</th>
<th>Feedback Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greeting</td>
<td>polite</td>
<td>smile, nod</td>
</tr>
<tr>
<td>Greeting</td>
<td>impolite</td>
<td>look sceptical, nod</td>
</tr>
<tr>
<td>Request value</td>
<td>polite</td>
<td>thinking gesture</td>
</tr>
</tbody>
</table>

*Table 4.1. Examples of feedback gesture assignments*

```prolog
listener_gesture(currentParallelAct!(dialActTypeAttr!greeting@datatype&
    addressee!politeness!"polite"&
    listenerFirst!["Blink1"] &
    listenerSecond!["Blink1"]&
    listenerThird!["Smile","Nod"]&
    listenerFourth!["Nod"]
  ,_TempOrd,_DialActs).
```

*Figure 4.2. Prolog code describing the predicate listener_gesture*

This code sample means basically that a polite person gives feedback to a DA of type *greeting* through smiling and nodding.

### 4.1.2 Parallel Dialogue Act Generator - ParallelGen

This module creates the Parallel Dialogue Acts. It contains a predicate called *createParallelDA* that takes care of the creation of the Parallel Dialogue Acts. A part of the predicate code is presented in Figure 4.3. The predicate arguments are:

- **DA** is a Dialogue Act.
- **ActionIndex** is a random number.
createParallelDA(DA,ActionIndex,RoleProb,TempOrd,DialActs,PDA):-
\+(DA=dialActTypeAttr\!openingComplaintResponse\! datatype) or
(DA=dialActTypeAttr\! feedback\! datatype) or
(DA=dialActTypeAttr\! refuse\!Answer\!Response\!datatype)),
actionDecision(DA,ActionIndex,RoleProb,ActionBool),
DA=dialActTypeAttr\! Type\!addresssee\!agreeableness\!Agree,
ActionBool=’true’,
DA=id!DaID,
  DA=speaker\!id!SpeakerId,
  DA=addresssee\! id\! ActorId,
  listener\_gesture(current\!Parallel\!Act\!DA &
        listenerFirst\! First &
        listenerSecond\! Second &
        listenerThird\! Third &
        listenerFourth\! Fourth,TempOrd,DialActs)

Figure 4.3. Prolog code describing the predicate createParallelDA

RoleProb is the probability that the role (seller or buyer) should do a feedback
gesture in this specific dialogue (more about this in section 4.1.3).

TempOrd is the Temporal Ordering element.

DialActs is the set of all DAs.

PDA is the output PDA.

The predicate first sorts out the DA:s with the “forbidden” DA types discussed in
section 3.2.2 (openingComplaintResponse, feedback or refuse\!Answer\!Response). After
that, the predicate actionDecision in FeedbackDecider (described in section 4.1.3) is
called to decide if this PDA shall contain a gesture. Finally, if the FeedbackDecider
gives green light, a gesture is chosen in feedbackGesticon (section 4.1.1) with the
predicate listener\_gesture. After this procedure the PDA is also added to the set
of Dialogue Acts and to the Temporal Ordering element. This happens in the main
module of the system and can not be seen in the code above.

4.1.3 Feedback Decider - FeedbackDecider

The module Feedback Decider contains the predicate that decides whether a feed-
back gesture is to be performed or not in a PDA. It contains two predicates. The
predicate setFeedbackProbabilities has the commission to calculate the probabilities
for the two roles in the dialogue (buyer and seller) to perform a feedback gesture.
This calculation is made only once per dialogue and is called in the main module
while the dialogue generation is started. This probability depends on how many “for-
bidden” Dialogue Acts there is in the dialogue. The algorithm is explicitly explained
in Appendix F.
The second predicate is called \textit{actionDecision} and a code sample from it is shown in Figure 4.4. The decision is randomly taken. \texttt{ActionIndex} is a random number between 1 and 100. The \texttt{RoleProb} is the role probability calculated by \textit{setFeedbackProbabilities}, and it is also a number between 1 and 100. The idea is that the buyer is more likely to give feedback and therefore the role probability of the buyer is higher than the role probability for the seller. The outcome, \texttt{ActionBool}, is true if the random number is less than the role probability and in that case a feedback gesture will be performed.

\textbf{4.2 Evaluation Results}

The results from the evaluation are the answers people gave on the questionnaire. There were three different types of questions and these were considered as basis for confirming or rejecting the hypotheses:

1. Audience find a dialogue \textit{with} feedback gestures more entertaining than a dialogue \textit{without} feedback gestures.

2. Humans find the behaviour of characters in a dialogue \textit{with} feedback gestures more natural than the behaviour of characters in a dialogue \textit{without} feedback gestures.

3. A dialogue \textit{with} feedback gestures is interpreted as more lively than a dialogue \textit{without} feedback gestures.

4. It is easier to remember the content of a dialogue \textit{with} feedback gestures than the content of one \textit{without} feedback gestures.

The answers to the first type of questions, the statements that the subjects were judging whether they agreed with or not, are considered in section 4.2.1. The second type tested the subjects’ abilities to remember details from the spoken dialogue and the results from this part is presented in section 4.2.2. Finally, various answers were given to the open-ended questions in the evaluation and a summation of these comments are shown in section 4.2.3.

Throughout the chapter the dialogue with feedback gestures is referred to as \textit{DFB+} and the dialogue without feedback gestures as \textit{DFB-}.

To measure the variables \textit{Entertainment (Ent)}, \textit{Naturalness of characters’ behaviour (Nat)} and \textit{Liveliness of dialogue (Liv)} the subjects were told to judge how
much the agreed with some statements. These statements together with the variables they are connected to are:

1. The dialogue was interesting to watch. (Ent)
2. The dialogue was entertaining. (Ent)
3. The dialogue made me laugh. (Ent)
4. Tina’s body language looked natural while she was speaking. (Nat)
5. Ritchie’s body language looked natural while he was speaking. (Nat)
6. Tina’s body language looked natural while she was NOT speaking. (Nat)
7. Ritchie’s body language looked natural while he was NOT speaking. (Nat)
8. The characters moved a lot while speaking. (Liv)
9. The characters moved a lot while NOT speaking. (Liv)

The judgement was made through putting a number on a scale from 1 to 5, where 1 means I totally disagree and 5 means I totally agree with the statement. We can directly measure the variables through combining the judgements from the related questions. The total mean value of the subjects’ judgements of the separate questions and the merged variable judgements are shown in Figure 4.5. Recall the course of events for the two evaluation sessions in Table 4.2. The two groups were shown the dialogues in different orders. In order to discover the effects the order of the dialogues had on the result we compare the differences between the two groups for $DFB^+$ and $DFB^-$ in Figure 4.6.
<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(DFB^+) was shown</td>
<td>(DFB^-) was shown</td>
</tr>
<tr>
<td>2</td>
<td>Questions, part 1</td>
<td>Questions, part 1</td>
</tr>
<tr>
<td>3</td>
<td>(DFB^-) was shown</td>
<td>(DFB^+) was shown</td>
</tr>
<tr>
<td>4</td>
<td>Questions, part 2</td>
<td>Questions, part 2</td>
</tr>
</tbody>
</table>

Table 4.2. Course of events in the Evaluation sessions

4.2.1 How Entertaining were the Dialogues?

First we look at the Entertainment variable (\(Ent\), statements 1-3). We see that the subjects found \(DFB^-\) more entertaining than \(DFB^+\). Looking specifically at the separate statements we see that the subjects found the two dialogues equally interesting but they laughed more at and found \(DFB^-\) more entertaining.

Comparing the two groups apprehension about \(DFB^-\) in Figure 4.6 we see that Group 2 (that saw this dialogue first) laughed more than Group 1 at \(DFB^-\) but Group 1 found that dialogue more interesting than Group 2 did. The both groups found the dialogue equally entertaining. According to \(DFB^+\), Group 1 was more positive on all points according to entertainment.

4.2.2 How Natural was the Characters’ Behaviour?

Let us now consider Naturalness (\(Nat\), statements 4-7). We return to Figure 4.5. The two diagrams show that subjects found the characters’ behaviour in \(DFB^-\) more natural than in \(DFB^+\). This is true for all statements connected to the Naturalness variable. Interesting is, that even the characters’ behaviour while not speaking (statement 6 and 7) were apprehended as more natural in \(DFB^-\).

The inter-group comparison in Figure 4.6 tells us that according to naturalness in \(DFB^-\), Group 1 gave more positive answers than Group 2. The two groups had the same apprehension about naturalness in \(DFB^+\). Group 1 was slightly more positive about the behaviour of the characters while listening in this dialogue.
4.2.3 How Lively were the Dialogues?

A final look at Figure 4.5 show us that when judging Liveliness (Liv, statements 8-9), the subjects seem to have noted that the listening characters moved more in DFB+ than in DFB-. The subjects apprehended the speakers’ movements more or less equally lively in DFB- and DFB+ only slightly livelier in DFB-.

Comparing the two groups in Figure 4.6 we see that for some reason Group 2 found the dialogues more lively in general except the listener movements in DFB-.

4.2.4 Results from Memory Testing

Next part of the evaluation was the memory test. The hypothesis was that subjects would remember the content of the DFB+ better than of the DFB-. The memory questions were only given once, after the first dialogue, i.e. Group 1 had only seen DFB+ then and Group 2 only DFB-. The questions were:

1. How much does the car cost?
2. How many miles per gallon does the car get?
3. Does the car have airbags?
4. Is the car environmental friendly?
5. Does the car have a large luggage compartment?
6. Does the car have leather seats?

The answers were given as statements from the car seller in the dialogue but some information needed to answer these questions was not present in the dialogue. The result of this memory test in terms of percent correct answers per group and question is presented in Figure 4.7. We see that Group 2 have 100 % correct answers to 5 of 6 questions and to the remaining one their result is still better than the result from Group 1.
4.2.5 Results from Open-ended Questions

Two open-ended questions were posed. Question 1 after the first dialogue and question 2 when both dialogues had been seen. Here, a summary of the answers from the two groups are given and the literal answers can be found in Appendix G.

Question 1, posed after the subjects had seen one dialogue. Group 1 had been shown $DFB^+$ and Group 2 had seen $DFB^-$.  

Did you think that the dialogue between Tina and Ritchie looked natural? Write down your apprehension about how natural their gestures were and why.

**Group 1 after seen $DFB^+$**. The subjects thought the dialogue was decent natural, but that the movements were too stiff and too robotic. The opinions differed about which character was most natural, someone said Tina and somebody else thought Ritchie looked more natural in his behaviour.

**Group 2 after seen $DFB^-$**. The general opinion seems to be that parts of the dialogue was natural but that the gestures were a bit jerky and exaggerated.

Question 2, posed at the very end of each evaluation session, asks for a subjective comparison of the two dialogues the subjects had seen.

How did you apprehend the gesticulation in this dialogue? Did you notice any differences between the two dialogues and in that case which differences?

**Group 1 after seen $DFB^+$ and $DFB^-$**. Many of the subjects say that they did not apprehend any big differences at all. In spite of this fact a number of them say that they believe that there was less movement in the latter dialogue, which is a correct observation.

**Group 2 after seen $DFB^-$ and $DFB^+$**. More gesticulation was apprehended in the latter dialogue by a majority of the subjects. A couple of the subjects could even specify the difference, that the characters were gesticulating when not speaking. None of the subjects in Group 1 made this observation. The subjects did not value this difference as “better” or “worse” it was more like a statement. One subject said that it made the dialogue look more connected. Many subjects thought that maybe they apprehended more movements because they paid more attention to the latter dialogue.

We can see that the subjects in general were not overwhelmed by the naturalness of the gestures in the dialogues. Many subjects found the gestures stiff and coarse, since we use ready made gestures from a collection of gestures for the MS Agents.
Chapter 5

Conclusions

In general, no big differences in the subjects’ impressions of the two dialogues, $DFB^+$ with feedback and $DFB^-$ without feedback, could be found from the evaluation. The judgements according to entertainment and naturalness were slightly more positive for the $DFB^-$ without gestures and the $DFB^+$ was judged as livelier. In this section some conclusions are drawn according to the hypotheses that the evaluation investigated. The evaluation results in section 4.2 are used as basis for these conclusions and they are discussed in order of the investigated variables. That is, entertainment (section 5.1), naturalness (section 5.2), liveliness (section 5.3) and memory (section 5.4). One general observation is that the differences in the given answers are small and therefore not statistically reliable. Another remark is that there were only six subjects in each group and that one can not draw general conclusions out of 12 persons’ judgements. At the end of this chapter, section 5.5 some conclusions drawn from the open-ended questions are described.

5.1 Entertainment

Recall the hypothesis “Audience find a dialogue with feedback gestures more entertaining than a dialogue without feedback gestures”. The results in section 4.2.1 show that the subjects judged $DFB^-$ as more entertaining than $DFB^+$. We can therefore conclude that the hypothesis must be rejected in this case. According to the comparison between the groups we see that both groups judged the first dialogue they were shown as more entertaining than the second dialogue. The dialogue seems to lose entertainment value second time subjects watch it. This is true independent of the presence of feedback gestures in the first dialogue. For all, the subjects laugh more at the first dialogue they see.

5.2 Naturalness

The hypothesis was that “Humans find the behaviour of characters in a dialogue with feedback gestures more natural than the behaviour of characters in a dialogue without
feedback gestures. When looking at the evaluation results we are forced to also reject this hypothesis in this case. There were four statements about naturalness for the subjects to judge. The subjects should specify how natural they found Ritchie’s and Tina’s behaviour while speaking and not speaking. In all cases the behaviour in $DFB_-$ was judged as more natural than in $DFB_+$. The characters’ behaviour in $DFB_-$ while not speaking was that they were not doing anything at all. In $DFB_+$ this behaviour was the feedback gestures that we are investigating the effects of. The subjects found the listeners’ behaviour more natural in $DFB_-$, when the characters did nothing, than in $DFB_+$, when the characters did feedback gestures. In the comments the subjects gave in the open-ended question many subjects say that the gesture movements are stiff and jerky. In this case a conclusion is then, that when there is more movements (e.g. in $DFB_+$) subjects seem to apprehend the whole dialogue as less natural because they think that the movements themselves are not natural. More movements mean then more unnatural movements and a less natural dialogue as a whole.

5.3 Liveliness

The third hypothesis was that “A dialogue with feedback gestures is interpreted as more lively than a dialogue without feedback gestures”. The results from the judgements of the statements about liveliness show that to some extent this hypothesis holds. The subjects agreed more with the statement “The characters moved a lot while not speaking” when they had seen $DFB_+$. Therefore we can conclude that the added feedback gestures were observed as additional liveliness. Liveliness and naturalness does not seem to be connected in this case. From the open-ended question we note that subjects in Group 2, that first were shown $DFB_-$ and then $DFB_+$, could specify the correct difference between the dialogues while subjects in Group 1 could not. Some subjects in Group 1 noted less movement in $DFB_-$ but they were not able to tell which kind of movement that was absent compared to $DFB_+$. This could be an effect of so called schemas. Subjects are humans and humans gesticulate when speaking. If the subjects find the characters enough human-like to expect them to act like humans they would fill in gestures when the characters do not gesticulate when reacting on speech as a human would have done. This This is the case for Group 1 that first saw $DFB_+$. It is made unwitting and therefore was it hard for the subjects to point out the difference between the two dialogues. The subjects in the other group, that first saw the dialogue without feedback gestures could specify the difference. The reason for this could be that they found the characters less human in the first dialogue and did not fill in the lacking feedback gestures in that case. When they then were shown the dialogue with feedback gestures they could, because the did not fill in the gestures before, observe the actual difference. Maybe this result give a hint that $DFB_+$ is more human-like than $DFB_-$ in some sense. It could also be that the schema was made from the first dialogue they were shown. The subjects that first saw $DFB_+$ expected the $DFB_-$ to contain feedback gestures
since the first dialogue did. When it did not, they completed the dialogue with the gestures themselves as described. The subjects in Group 2 made a schema of the dialogue with no feedback gestures. When the second dialogue contained feedback gestures they noted it and could also specify the difference.

5.4 Memory

Through testing the subjects' abilities to remember some details from the first dialogue they were shown the following hypothesis was tested: "It is easier to remember the content of a dialogue with feedback gestures than the content of one without feedback gestures". Unfortunately the subjects in Group 2 that had seen DFB- were better at remembering the details than the subjects in Group 1. We must therefore reject also the hypothesis about the presence of feedback gestures has a positive influence on subjects' abilities to remember the dialogue content. The subjects watching DFB+ might have been less observant to the spoken details in the dialogue since there are more movements to pay attention to.
Chapter 6

Discussion and Future work

Here follow some personal reflections of the project. First I discuss the problem formulation and the method I used to complete the task. Then I discuss the conclusions and give some suggestions of future work on the subject.

6.1 Discussion on Chosen Task and Method

The original task

*Given the content of a Scene Description, how can we realize feedback gestures and which influence does the presence of feedback gestures have on the total impression of the dialogue according to entertainment, naturalness of characters' behaviour and liveliness?*

was completed through designing and implementing feedback gestures into the NECA system and then evaluating the modified system. One disadvantage with the task was that some gestures already were implemented in NECA and this implementation had a restraining effect on my creativity to find a new solution for the generation of the feedback gestures I was implementing. Looking back, I wish I had focused more on the abstract representation of the gestures. Instead the generation turned out to be a direct mapping to the limited collection of gestures that is available in the MSAgent used in NECA. On the other hand, the fact that adding feedback gestures would really change the appearance of the dialogue as a whole was exciting and stimulating.

The method of implementation by representing the feedback gestures in separate parallel dialogue acts had some advantages. The implementation phase was simplified by this representation since very little of the earlier code needed to be changed, new modules were written and integrated. A disadvantage might be that the meaning of a dialogue act was changed a bit. It does no longer contain everything that happens at a specific moment in the dialogue but only what the speaker is doing. On the other hand the DA together with the PDA have all this information and since the PDA connected to the DA easily can be found this should be clear enough.
The evaluation suffered from a lack of participants. The results can therefore not be taken as absolute facts but rather as a hint of how humans apprehend the dialogues. It was good to use two separate groups to eliminate the effects of the order in which the dialogues were shown. The variables were hard to define and to figure out how to measure. Maybe the statements were too direct and maybe more open-ended questions to let the subjects themselves formulate their impressions of the dialogues would have been a good thing. More statements connected to each variable would also have given a statistically more solid result. Regarding the memory test, a pre-memory test should have been made on the two groups to state their memory abilities compared to each other first. This could have been done by letting them see a picture and then describe it.

6.2 The Relevance of the Conclusions and Future Work

The conclusions must not, as earlier mentioned, be taken as absolute facts but rather as indications of humans’ reactions on the NECA dialogues. The aim of the whole dialogue is to give the audience some information in a entertaining way. The person watching the dialogue should not necessarily think that the dialogue is very natural but on the other hand it should not be so stiff and unnatural that the person loses interest in the dialogue. I think that it is important to regard the fact that many of the subjects found the gestures themselves stiff and jerky and that this had an effect on the total impression of the dialogue. Maybe the most important is not the quantity but the quality of the movements of embodied agents. I suggest a small but natural amount of gestures in this kind if dialogue. If the characters achieve a more natural appearance with for example breathing movements, one could also an extension of the amount of gestures make the dialogue look more human resemble.

It is also an interesting psychological effect that subjects that had seen the $DFB^+$ with feedback gestures first, could not specify the difference between the two dialogues but some of the subjects that had seen $DFB$ without feedback first, could. This kind of experiment would be interesting to do with a larger group. The memory test could also be re-made with some kind of memory calibration as described in the previous section.
Chapter 7

References

Allwood,J. 1999 The Swedish spoken language Corpus at Göteborg University, Fonetik 99, Gothenburg Papers in Theoretical Linguistics 81


Cerrato,L. 2002 A comparison between feedback strategies in Human-to-Human and Human-Machine communication, in proceedings of ICSLP-2002, Denver, Colorado USA


Jurafsky,D. and Martin,J.H. 2000 Speech and Language Processing, Prentice Hall, 2000 pp 765-768


Appendix A

Gesture generation in NECA

Here, the complete gesture generation in NECA at the starting point of the project is described. The generation is based on DA type and the structure is that the gestures fill different functions in combination with different DA types.
<table>
<thead>
<tr>
<th>Functions &amp; DA values</th>
<th>Communicative behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function: Greeting</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= greeting and politeness= polite and speaker= Tina,</td>
<td>Block gesture and look at addressee</td>
</tr>
<tr>
<td>DA type= greeting and politeness= impolite</td>
<td>Fold gesture and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Question</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= request if or request value</td>
<td>Hands on the hips or raise hand and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Inform</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= inform</td>
<td>Cup one hand and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Opening complaint</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= opening complaint</td>
<td>Look around</td>
</tr>
<tr>
<td><strong>Function: Negative response</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= negative response</td>
<td>look sceptical and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Opening response or request information</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= opening response or ‘request info’</td>
<td>Look at the car and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Refuse answer</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= refuse answer</td>
<td>Shrug shoulders and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Feedback</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= feedback</td>
<td>Smile and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Negative evaluation</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= negative evaluation</td>
<td>Lower both hands and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Disconfirm</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= disconfirm</td>
<td>Raise eyebrows and look at addressee</td>
</tr>
<tr>
<td><strong>Function: Complete closing</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= complete closing and politeness= impolite</td>
<td>Laugh</td>
</tr>
<tr>
<td><strong>Function: Initiate closing</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= initiate closing positive and speaker= Tina</td>
<td>Touch face thinking and look at addressee</td>
</tr>
<tr>
<td>DA type= initiate closing positive and speaker= Ritchie</td>
<td>cover eyes with hand and look at addressee</td>
</tr>
<tr>
<td>DA type= initiate closing negative and speaker= Tina</td>
<td>throw out arms and look at addressee</td>
</tr>
<tr>
<td>DA type= initiate closing negative and speaker= Ritchie</td>
<td>cover eyes with hand and look at addressee</td>
</tr>
</tbody>
</table>

**Table A.1.** Gesture generation in NECA at starting point
Appendix B

Feedback gesture generation in NECA

Here, the complete feedback gesture generation in NECA at the end of the project is described. The generation is based on DA type and personality values and the structure is similar to the original gesture generation.

<table>
<thead>
<tr>
<th>Feedback functions &amp; original DA types</th>
<th>Feedback communicative behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function: Greeting</strong></td>
<td></td>
</tr>
<tr>
<td>DA type = greeting and politeness = polite</td>
<td>Smile or nod and nod</td>
</tr>
<tr>
<td>DA type = greeting and politeness = impolite</td>
<td>Look sceptical or nod</td>
</tr>
<tr>
<td><strong>Function: Acknowledge comprehension</strong></td>
<td></td>
</tr>
<tr>
<td>DA type = request if or request value or request info</td>
<td>Touch face thinking or fold gesture</td>
</tr>
<tr>
<td>DA type = opening question and politeness = impolite and listener = Tina</td>
<td>Put on glasses or fold gesture</td>
</tr>
<tr>
<td>DA type = opening question and politeness = polite and listener = Tina</td>
<td>Smile or nod</td>
</tr>
<tr>
<td>DA type = opening question and listener = Ritchie</td>
<td>Smile and nod</td>
</tr>
<tr>
<td>DA type = opening complaint and neuroticism = neurotic and listener = Tina</td>
<td>Look nervous</td>
</tr>
<tr>
<td>DA type = opening complaint and neuroticism = not neurotic and listener = Tina</td>
<td>Smile</td>
</tr>
<tr>
<td>DA type = opening complaint and listener = Ritchie</td>
<td>Smile and fold gesture</td>
</tr>
<tr>
<td>Feedback functions &amp; original DA types</td>
<td>Feedback communicative behaviour</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><strong>Function: Show continuation</strong></td>
<td></td>
</tr>
<tr>
<td>DA type = inform and politeness =</td>
<td>progress gesture</td>
</tr>
<tr>
<td>impolite and next DA type is also</td>
<td>inform with the same speaker.</td>
</tr>
<tr>
<td><strong>Function: Express a point of view</strong></td>
<td></td>
</tr>
<tr>
<td><strong>and acknowledge comprehension</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= inform and politeness =</td>
<td>Smile and nod</td>
</tr>
<tr>
<td>polite</td>
<td></td>
</tr>
<tr>
<td>DA type= inform and politeness =</td>
<td>Look sceptical or nod or put hands on hips and fold gesture</td>
</tr>
<tr>
<td>impolite</td>
<td></td>
</tr>
<tr>
<td>DA type= disconfirm or refuse answer</td>
<td>Fold gesture or hands on hips or look sceptical</td>
</tr>
<tr>
<td>DA type= opening response and politeness = impolite and listener =</td>
<td>Put on glasses</td>
</tr>
<tr>
<td>Tina</td>
<td></td>
</tr>
<tr>
<td>DA type= opening response and politeness = polite and listener = Tina</td>
<td>Smile or nod</td>
</tr>
<tr>
<td>DA type= opening response and listener = Ritchie</td>
<td>Nod or smile and fold gesture</td>
</tr>
<tr>
<td>DA type= negative response and</td>
<td>Hands on hips and fold gesture</td>
</tr>
<tr>
<td>politeness = impolite and agreeableness = disagreeable</td>
<td></td>
</tr>
<tr>
<td>DA type= negative response and</td>
<td>Smile and nod</td>
</tr>
<tr>
<td>politeness = impolite and agreeableness = agreeable</td>
<td></td>
</tr>
<tr>
<td>DA type= negative response and</td>
<td>Smile and nod</td>
</tr>
<tr>
<td>politeness = polite</td>
<td></td>
</tr>
<tr>
<td><strong>Function: Closing</strong></td>
<td></td>
</tr>
<tr>
<td>DA type= complete closing</td>
<td>Smile and nod</td>
</tr>
<tr>
<td>DA type= initiate closing positive</td>
<td>Nod</td>
</tr>
<tr>
<td>and politeness = impolite</td>
<td></td>
</tr>
<tr>
<td>DA type= initiate closing positive</td>
<td>Smile</td>
</tr>
<tr>
<td>and politeness = polite</td>
<td></td>
</tr>
<tr>
<td>DA type = initiate closing negative</td>
<td>Shake the body</td>
</tr>
<tr>
<td>DA type= complete closing negative</td>
<td>Fold gesture or hands on hips or look sceptical</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.1. Feedback gesture generation in NECA
Appendix C

Flow Charts

This appendix contains two flow charts. Flow chart 1 is a model of the information flow for the gesture generation in the NECA system at starting point. Flow chart 2 was created in the designing phase of the project and shows which new modules that had to be created and how they were supposed to be integrated with the original NECA system.

Figure C.1. Flow chart over information flow in NECA gesture generation at the starting point
Figure C.2. Gesture and feedback gesture generation in NECA. Design flow chart
Appendix D

Questionnaire

This appendix contains the questionnaire that was handed out to the subjects in the two evaluation sessions. It contains both the original version in Swedish and a translation of it in English.
Frågeformulär

Instruktioner

Mitt namn är Malin Bergenstråhle och jag hälsar dig hjärtligt välkommen till detta lilla test som är en del av mitt examensarbete. Du behöver inte ange ditt namn, utan bara kön, ålder och sysselsättning. Inga uppgifter kommer att utlämnas så att de kan förknippas med dig som person.

För den första typen av frågor finns fem cirklar som motsvarar en skala från 1 till 5 mellan några extremvärden som visas i samband med frågan på skärmen.
Frågorna av den andra typen svarar du på genom att skriva ner ditt svar vid den siffra som frågan har.

Frågorna visas i några sekunder och du behöver inte tänka efter så mycket, utan sätt bara ett kryss för det som spontant känns riktigt.

-----------------------------
Frågor att svara på innan testet

Kön

Man  O  Kvinna  O

Ålder  _____ år

Sysselsättning  ____________________________________________________________

__________________________________________________________________________
Frågor efter första dialogen

<table>
<thead>
<tr>
<th>Fråga 1.1:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fråga 1.2:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Fråga 1.3:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Fråga 1.4:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Fråga 1.5:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Fråga 1.6:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

| Fråga 1.7: | 1 | 2 | 3 | 4 | 5 |
| Fråga 1.8: | 1 | 2 | 3 | 4 | 5 |
| Fråga 1.9: | 1 | 2 | 3 | 4 | 5 |

<table>
<thead>
<tr>
<th>Fråga 1.10:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fråga 1.11:</td>
<td></td>
</tr>
<tr>
<td>Fråga 1.12:</td>
<td></td>
</tr>
<tr>
<td>Fråga 1.13:</td>
<td></td>
</tr>
<tr>
<td>Fråga 1.14:</td>
<td></td>
</tr>
<tr>
<td>Fråga 1.15:</td>
<td></td>
</tr>
</tbody>
</table>

Svar:

---

---
Frågor efter andra dialogen

Fråga 2.1:  
1 2 3 4 5

Fråga 2.2:  
1 2 3 4 5

Fråga 2.3:  
1 2 3 4 5

Fråga 2.4:  
1 2 3 4 5

Fråga 2.5:  
1 2 3 4 5

Fråga 2.6:  
1 2 3 4 5

Fråga 2.7:  
1 2 3 4 5

Fråga 2.8:  
1 2 3 4 5

Fråga 2.9:  
1 2 3 4 5

Fråga 2.10:  
Svar:  

Tusen tack för din medverkan!  
Malin Bergström
Questionnaire

Instructions

My name is Malin Bergenstråhle and You are warmly welcomed to this small test that is a part of my Master’s Project. You do not have to state your name, only sex, age and occupation. No personal information will be exposed so that they can be associated with Your person.

You will be shown 2 dialogues between a couple of animated characters named Ritchie and Tina. After each dialogue a number of questions will be shown on the screen, one at a time. You are supposed to give Your answers to the questions on the following ansering sheets. The questions are of two different kinds.

To the first type of question there are five circles to each question, corresponding a scale from 1 to 5 between some extreme values that are shown together with the question on the screen. You give Your answers to the second type of questions through writing Your answer beside the number of the question.

The questions are shown a few seconds and You do not have to ponder much, just put a cross in the circle that spontaneously feels correct.

Questions to be answered before the test

Sex  
- Male  
- Female

Age  

years

Occupation  


### Questions after the first dialogue

<table>
<thead>
<tr>
<th>Question 1.1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.10:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.11:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.12:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.13:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.14:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 1.15:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questions after the second dialogue

Question 2.1:
1 = totally disagree  2  3  4  5 = totally agree

Question 2.2:
1  2  3  4  5

Question 2.3:
1  2  3  4  5

Question 2.4:
1  2  3  4  5

Question 2.5:
1  2  3  4  5

Question 2.6:
1  2  3  4  5

Question 2.7:
1  2  3  4  5

Question 2.8:
1  2  3  4  5

Question 2.9:
1  2  3  4  5

Question 2.10:
Answer:

Many thanks for Your participation!
Malin Bergstråhle
Appendix E

Software documentation

I have added and changed some module in the NECA system so that for each Dialogue Act (DA) a parallel DA is created. This parallel Dialogue Act contains information about what the agent listening is supposed to do during the original DA. The listener might do some feedback gesture or just stand still.

This documentation contains descriptions of the role of each added module and a specific description of their predicates visible to other modules.

E.1 Added Modules

Here I describe the new modules that were added to the NECA system. The predicates and their in- and outputs are defined and their functions described.

Module FeedbackDecider

Location: m_nlg/src/appl/eshowroom/data/FeedbackDecider

Available Prolog library:

- lists

This module is application dependent and contains two predicates to decide whether the listener to a specific DA should do a gesture or not.

Predicates

setFeedbackProbabilities(DialActs,SellerProb,BuyerProb) this predicate is called directly when the program is started. It takes as input the set of original Dialogue Acts and returns the probabilities for the seller and the buyer to do a feedback gesture. The idea behind these probabilities origins from two assumptions. The first is that the buyer is more likely to give feedback than the seller (Cerrato, 2002) and the second is that some DA types are not likely at all to give feedback to. These types are

- refuseAnswerResponse
• openingComplaintResponse
• feedback

and the reason is that these DA types are feedback already and it might not be likely to give feedback on feedback. The algorithm contains an assumed probability for the seller and an assumed probability for the buyer to do a feedback gesture. It then calculates the number of DAs of the types that never should have feedback gestures and then changes the probabilities for the remaining DAs so that the correct total probability is preserved. A more explicit description of this algorithm is available in Appendix E.

actionDecision(DA, ActionIndex, RoleProb, ActionBoolean) this predicate is called from the createParallelDA predicate in the ParallelGen module. For each parallel DA (PDA) that is created there is a randomized choice whether it shall contain a feedback gesture or not. If the original DA type is one of the “forbidden” types mentioned above the PDA directly gets a label “wait” that indicates that the listener shall not move during this DA. Else a random number, ActionIndex, between 1 and 100 is compared to 100*the probability for the role that the listener has (buyer or seller) to do a feedback gesture, Random <= Probability. The result is converted to a boolean output, ActionBoolean. If, for example the probability for the buyer to do a feedback gesture is 0.6 and the random number is 43, the comparison 43 <= 60 is true, ActionBoolean gets the value “true” and the listener will do a feedback gesture.

Module ParallelGen

Location: m_nlg/src/system/modules/ParallelGen
Available Prolog libraries:
• lists
• random
Available modules:
• m_nlg/src/system/lib/listsUtil
• m_nlg/src/system/lib/counterUtil
• m_nlg/src/system/lib/randomUtil
• m_nlg/config/config
• m_nlg/src/appl/eShowroom/data/feedbackgesticon
• m_nlg/src/appl/eShowroom/data/FeedbackDecider
• m_nlg/src/appl/eShowroom/data/MSAgentMapping

53
This module is application independent and its single visible predicate createParallelDA/6 takes care of the creation of the parallel Dialogue Acts (PDA). It is called for each DA in realizeDialogueAct in the main module. The returned PDA is then added to the Set of Dialogue Acts and to the temporal order.

Predicate

createParallelDA(DA, ActionIndex, RoleProb, TempOrd, DialActs, PDA) this predicate is called in Main. Output is the parallel DA, PDA. To call this predicate one must have a random number ActionIndex, here generated by the built-in random number generator in Prolog. One must also have a probability that this PDA shall contain a feedback gesture. This probability is role dependent (eg, different for the buyer and seller) and was calculated by setFeedbackProbabilities in the module FeedbackDecider when the program was started. The predicate then uses actionDecision to decide whether the PDA will contain a feedback gesture or not. Finally, if actionDecision decides that a gesture shall be performed, the predicate calls listener_gesture in module feedbackGesticon to decide which gesture is suitable.

Module feedbackGesticon

Location: m_nlg/src/appl/eShowroom/data/feedbackGesticon
Available Prolog library:

- lists

This module is application dependent and its function is similar to module gesticon that chooses speaker gestures, while feedbackGesticon chooses listener gestures. Its single predicate listener_gesture checks the DA type and some of the personality values such as politeness and in some case also agreeableness and neurosis.

Predicates

listener_gesture(DA, TempOrd, DialActs) this predicate is called by createParallelDA in module ParallelGen. It checks the DA type and some other parameters and assigns a feedback gesture to the PDA. For each combination of DA type and personality values it has one or more feedback gestures and it chooses one by random choice. The temporal order and the set of dialogue acts are also taken as arguments only because of one situation: If the current DA is an Inform and the next also is an Inform with the same speaker, then the listener has the possibility to do a “show continuation” feedback gesture.
E.2 Changed Modules

Module MSAgentGen

I totally changed this module so that it now takes a RRL document and makes a MSAgent script out of that information instead of making it directly from the Prolog list of sentences and gestures as it did before.
Appendix F

Algorithm for calculating probabilities

This Appendix consists of a description of the algorithm that re-culates the probability for the roles to give feedback. The total probability for the buyer and the seller should remain the same, the task is to control the number of forbidden DA types in each generation of a Dialogue and then recalculate the probabilities.

Input: Set of dialogue acts
Output: Probability to be sent into the random action decision

Since there are three dialogue act types that should not have feedback gestures these are to be excluded at a stage before the random action decision is made. These types are:

- Feedback (speaker is buyer)
- OpeningComplaintResponse (speaker is seller)
- RefuseAnswerResponse (speaker is buyer)

Let us assume:
- $T_{NB}$=Total number of dialogue acts to which the buyer is to give feedback
- $T_{NS}$=Total number of dialogue acts to which the seller is to give feedback
- $No_{A_{buyer}}$=Number of dialogue acts that should have parallel feedback acts for the buyer, i.e. 60 % of $T_{NB}$ (seller: 30 % of corresponding $T_{NS}$)
- $No_{F_{buyer}}$= Number of “forbidden” dialogue acts, i.e. dialogue acts of the three types mentioned above.

Since we still want the same absolute number of dialogue acts to have parallel feedback acts we then have to change the probability so that in the buyer’s case

$$No_{A_{buyer}} = 0.6 \times T_{NB} = NewProb \times (T_{NB} - No_{F_{buyer}})$$

$$\Rightarrow NewProb = \frac{0.6 \times T_{NB}}{T_{NB} - No_{F_{buyer}}}$$

And this is the output probability. The procedure for calculating the seller’s new probability is analogous but with 0.3 instead of 0.6 and of course $No_{F_{seller}}$ and $T_{NS}$. The boundary situations are no forbidden DA types found all DA are of forbidden types
1. No forbidden types found. The $NoF_{\text{buyer}}$ is 0 and the $NewProb$ is 60 % as before.

2. All DAs are forbidden. Since the denominator in the fraction then becomes 0 and the value of the fraction is infinitely we have to make sure that this does not happen. We avoid this situation through the following control: If $NoF_{\text{buyer}}$ exceed the 40% of $TNB$ that was predefined, we simply set the probability to 100 % for the remaining dialogue acts to have action, since a probability over 100 %

An example: Imagine 10 Dialogue Acts that the buyer possibly could give feedback to. We want 6 of them to have feedback and 4 of them not. 2 of them are “forbidden”. Now, we still want that 6 Dialogue Acts should have feedback but this time it is 6 out of the 8 remaining Dialogue Acts. We calculate the new probability with the formula and found it to be 75 %, since $0.75 \times 8 = 0.6 \times 10 = 6$. 

57
Appendix G

Answers to open-ended questions

This appendix contains translations of the literal answers that the subjects in the evaluation sessions gave to the two open-ended questions that were posed.

Group 1

Question 1, after seen DFB+

Did you think that the dialogue between Tina and Ritchie looked natural? Write down your apprehension about how natural their gestures were and why.

1. “They talked a lot with their hands, which looked natural.”

2. “Ritchie looked stiff and unnatural, he did not move much. Tina’s movements were contradictory. She bought the car despite her body language was “negative”. She looked more natural than Ritchie.”

3. “A bit jerky, but quite natural to be animated figures.”

4. “No, the gestures were maybe decent but the fact that the voice was totally indifferent and did not follow the content made the character of the conversation stiff.”

5. “I think there were too many robotic gestures.”

6. “Ritchie looked more natural, maybe because he moved more. Tina shook her head as if she said something negative, which was strange. But Tina turned her head when she was listening, which was good. Ritchie was funny, really a salesman with the hand and the arm.”
Question 2, after seen DFB+ and DFB-

How did you apprehend the gesticulation in this dialogue? Did you notice any differences between the two dialogues and in that case which differences?

1. “My experience was that the latter dialogue was less gesticulating.”

2. “It was less jerky than the first dialogue. It looked a bit more natural. I do not understand why Tina looked so ironic. Softer and more brief movements than the first dialogue.”

3. “When Tina was thinking about if she should by the car, she made a new gesture. Otherwise I did not notice any differences.”

4. “Roughly the same dialogue as the first, maybe they moved a bit less but otherwise I did not notice any differences.”

5. “Not much! The movements were still robotic. Furthermore, too long pauses between the two characters which is unnatural. “Speed”?”

6. “Tina was more flirtatious. She smiled and moved more. I think Ritchie was stiffer.”

Group 2

Question 1, after seen DFB-

Did you think that the dialogue between Tina and Ritchie looked natural? Write down your apprehension about how natural their gestures were and why.

1. “Natural gestures, but also a bit stiff and jerky.”

2. “The gestures were a bit theatrical, I suggest more modest movements if the dialogue is to be used in Sweden/Scandinavia/UK.”

3. “Some parts were natural but when they changed gesture was the change itself not very natural. It looked cut.”

4. “The gestures looked jerky and coarse probably is a higher level of details needed to make it look natural.”

5. “Exaggerated gestures, too long pauses, no “flow” in the dialogue.”

6. “Fairly natural, for all that they reacted on what the other person said was natural.”
Question 2, after seen DFB- and DFB+

*How did you apprehend the gesticulation in this dialogue? Did you notice any differences between the two dialogues and in that case which differences?*

1. “I remembered it better than it now appeared to be. This time I saw how unnatural it was because I watched the dialogue more attentive.”

2. “This time I apprehended a lot of gesticulation (partially because I looked for it more actively this time). Still too sweeping gestures to be real, for all Tina’s thinking expression.”

3. “Yes, I think they did more things when not speaking, more blinks and some arm movements. Did not apprehend any details but it looked “softer” and more connected.”

4. “Ritchie had a clearer body language. He stood with his arms crossed when not speaking. Tina’s right arm was moving more in the latter dialogue.”

5. “The difference was when they were not speaking. Tina gesticulated more when Ritchie was speaking, I think.”

6. “I apprehended almost no difference, possibly that the characters were gesticulating more in the latter dialogue. More head movements? Or maybe I just watched more carefully this time.”