



HAL
open science

Multimodal Intercultural Information and Communication Technology - A Framework for Designing and Evaluating Multimodal Intercultural Communicators

Jens Allwood, Elisabeth Ahlsén

► **To cite this version:**

Jens Allwood, Elisabeth Ahlsén. Multimodal Intercultural Information and Communication Technology - A Framework for Designing and Evaluating Multimodal Intercultural Communicators. Kipp, M., Martin, J-C., Paggio, P., & Heylen, D. Multimodal corpora: from models of natural interaction to systems and applications, Springer, pp.160-175, 2009, Lectures Notes in Computer Science; vol. 5509, 10.1007/978-3-642-04793-0_10 . hprints-00504104

HAL Id: hprints-00504104

<https://hal-hprints.archives-ouvertes.fr/hprints-00504104v1>

Submitted on 19 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This is an author produced version of a **contribution to**:

Kipp, M., Martin, J-C., Paggio, P., & Heylen, D. (eds) (2009). *Multimodal corpora: from models of natural interaction to systems and applications*. Berlin, Springer. (Lecture Notes In Computer Science; Vol. 5509)

This contribution has been peer-reviewed but may not include the final publisher proof-corrections or pagination.

Citation for the published contribution:

Allwood, J., & Ahlsén, E. (2009). Multimodal Intercultural Information and Communication Technology - A Framework for Designing and Evaluating Multimodal Intercultural Communicators. In: Kipp, M. et al (eds) *Multimodal corpora: from models of natural interaction to systems and applications*. (Lecture Notes In Computer Science; Vol. 5509). p 160-175.

Multimodal Intercultural Information and Communication Technology – A framework for designing and evaluating Multimodal Intercultural Communicators

Jens Allwood¹ and Elisabeth Ahlsén¹

¹ SSKKII Interdisciplinary Center, University of Gothenburg

Abstract. The paper presents a framework, combined with a checklist for designing and evaluating multimodal, intercultural ICT, especially when embodied artificial communicators are used as front ends for data bases, as digital assistants, as tutors in pedagogical programs or players in games etc. Such a framework is of increasing interest, since the use of ICT across cultural boundaries in combination with the use of ICT by persons with low literacy skills is rapidly increasing. This development presents new challenges for intercultural ICT. A desideratum for interculturally sensitive artificial communicators is a generic, exportable system for interactive communication with a number of parameters that can be set to capture intercultural variation in communication. This means a system for a Generic, Multimodal, Intercultural Communicator (a GMIC).

Keywords: multimodal ICT, intercultural ICT, virtual communicator, ECA (embodied communicative agent)

1 Purpose

This paper presents a framework, combined with a checklist, for designing and evaluating multimodal intercultural ICT (MMIICT). After motivating the study of MMIICT and defining the concept, the paper focuses on how a GMIC can be designed and/or evaluated with respect to adaptation to variation in activity and culture, using the checklist. Finally, an illustrating example of an evaluation of a web based embodied communicative agent (ECA) used in many countries is given.

2 Why multimodal intercultural ICT is an area of increasing importance

The use of ICT to support communication and information transfer across national, ethnic, cultural boundaries is becoming more and more common. Intercultural ICT, in this sense, can be present in intercultural use of e-mail, chat, digital news broadcasts,

blogs, games, intercultural education and multimodal websites. Especially interesting here is the use of multimodal agents, avatars and robots to communicate and give information across cultural boundaries. The use of such devices as front ends of data bases, in games and chat fora (Life World etc) is quickly increasing.

It is likely that this use will increase even more as people with low literacy skills become users of ICT, since this will be the most natural way for them to communicate. In this situation, it will become more and more interesting to have avatars and other ECA:s who possess natural (human like) communication skills. This development points to an increased need for ECA:s that can be adapted for use in different cultures and different activities of these cultures.

3 Definition of Multimodal Intercultural ICT

By “Multimodal Intercultural ICT”, we mean ICT which employs a multimodal GUI. Such a GUI uses two or more of the visual, auditory, tactile, olfactory and gustatory sensory modalities. It also uses two or more of the Peircean modes of representation (index, icon and symbol) [1]. Our focus will be on dynamic, interactive ICT employing avatars or other artificial communicators, across national, ethnic, cultural boundaries. We characterize an “avatar” as a VR representation of a user and an “artificial communicator” as any communicative agent with a multimodal or multirepresentational front end (cf. above). An avatar will in this way be a special case of an “artificial communicator” or “embodied communicative agent” (ECA).

4 Activity dependence of ICT

Both in design and evaluation, it is important to relate ICT to the social activity it is supposed to be a part of. Thus, there are different activity requirements if we compare an “artificial communicator” that has been constructed as a front end to a data base (e.g. for a multinational company to present its products), as a personal digital assistant, as a friendly tutor teaching small children to read and write or as an avatar which is to represent a player in a game like War Craft.

Everywhere the social activity, with its purpose, its typical roles, its typical instruments, aids, procedures and environment, determines what are useful characteristics of the “artificial communicator”. Both in designing a specification and in designing an evaluation schema, it is therefore important to build in systematic ways of taking activity dependence into account [2].

5 Generic applicability and multimodal robustness

A second desideratum for interculturally sensitive artificial communicators is to base them on a generic system for interactive communication with a number of parameters that can be set to capture intercultural variation in communication. For

interesting suggestions in this direction, see [3], [4]. Kenny et al. [3] focus on Virtual Humans used for training leadership, negotiation, cultural awareness and interviewing skills. Their goal is to create engaging characters that convey the three main characteristics of being *believable* (giving the illusion of human-like behavior), *responsive* (to the human user and the surrounding events, by having a rich inner dynamic) and *interpretable* (using the same “verbal and nonverbal cues that people use to understand one another”). They also distinguish three layers in a Virtual Human Agent: *the cognitive layer*, which “makes decisions, based on input, goals and desired behavior”, *the virtual human layer*, or body, including input processing (e.g. vision, speech, smell) and output processing (verbal speech, body gestures and actions) and *the simulation layer* (environment). Further, Kenny et al. point to the role of emotions in recognition and expression. This is also stressed by Kopp et al. [5].

A model presented by Jan et al. [4] provides parameters for different cultures (North American English, Mexican Spanish and Arabic) for a chosen subset of conversational behavior: *proxemics*, *gaze* and *overlap in turn taking*. Their scenario is also Virtual Humans in environments used mainly for training intercultural communication. They advocate a modular design where functional elements can be mapped to culture-specific surface behaviors. This has been done, for example, in the ECA GRETA [6].

There are not very many studies of the effects of cultural variation in avatars. Koda studied Japanese designed avatars in different Asian countries in and, in a follow-up study, western designed avatars also in North and South America. He found that there are cultural differences in how facial expressions are interpreted and that gestures could interfere with the interpretation of facial expressions [7], [8]. Koda and coworkers found a wide variation in the interpretation of positive expressions, whereas negative expressions were recognized more accurately [9]. Based on claims about cultural differences in the perception of avatars, Johansen [10] compared avatar perception by American and German users. The hypotheses were that American users, coming from an image dominated culture [11], would be more sensitive to attractiveness in an avatar, while German users would place more importance on credibility [12]. The study was also based on claims by Barber and Barde [13] and Chau et al. [14] about cultural differences in reactions to stimuli, for example, that a global interface has to be localized or designed according the cultural nuances of the target audience in order to be effective. Johnson’s results did show that Germans reacted more positively to a credible avatar than Americans, but in general similarities between the two groups were greater than expected. A generic multimodal intercultural communicator (GMIC), thus, has to be flexible and easily adapted to similarities as well as differences between different cultures and different activities.

Constructing a GMIC would mean constructing a generic system that in principle would allow similar contents or functions to be localized in a culturally sensitive manner which often might mean slightly different ways. It is important here to say “similar”, since the contents (e.g. news casts) or functions (e.g. giving advice) could themselves be affected by cultural variation [15]. Below, we will provide a suggestion (in the form of a kind of check list) for some of the parameters that could characterize such a system.

A third desideratum for the system is “multimodal robustness” in the sense that the system should be able to handle difficulties in text understanding, difficulties in speech recognition and difficulties in picture/gesture recognition in a sensible way. The system should not halt or respond by “unknown input” or “syntax error” each time routines for recognition or understanding break down. The GMIC should provide routines for how, given a particular activity, such problems can be handled, e.g. by being able to record user contributions, even if they are not recognized or understood and then playing them back to the user as a repetition with question intonation, or by giving minimal feedback through head movements or minimal vocal contributions (which have the function of encouraging the user to continue).

6 Some intercultural parameters of a GMIC

Below, we will present a number of features of communication, which exhibit cultural variation. The features are based on earlier work [20],[15], [3], [4].

6.1 Cultural variation in expressive behavior

Some expressive communicative behavior exhibits large scale cultural variation [16]. Besides verbal parameters, a GMIC needs to have parameters for

- head movements (nods, shakes, backward jerks, left turn, right turn, forward movement, backward movement)
- facial gestures (smiles, frowns, wrinkle, mouth movements other than speech)
- eye movements and gaze
- eye brow movements
- posture shifts
- arm and hand movements
- shoulder movements
- intonation in speech
- intensity, pitch and duration in speech

In all of these parameters [17] several fairly well attested (stereotypical) cultural differences exist, e.g. head movements for “yes” vary between typical European-style nodding and the Indian sideways wagging. Similarly, head movements for “no” vary between headshakes and the backward jerk with an eye-brow raise (sometimes called “head toss”), which is common from the Balkans through the Middle East to India [18], [19].

6.2 Cultural variation in content and function

Expressive behavior does not exist for its own sake, but in order to convey content. National, ethnic cultures vary in what expressions, content and functions are seen as allowable and appropriate in different contexts [20]. Should we always smile to strangers? Should women smile to men? Should voices always be subdued and

modulated or only when talking to people with higher status? How permissible are white lies? What is worse, a lying system or an insulting system?

Below are some content areas, where studies have shown cultural variation [16].

- Emotions. What emotions are acceptable and appropriate in different activities? E.g. is it permissible for two colleagues at work to quarrel and show aggression or is this something that should be avoided at all costs?
- Attitudes. What attitudes, e.g. regarding politeness and respect, are appropriate? Should titles and formal pronouns, rather than first names and informal pronouns be used?
- Everyday topics. What topics are regarded as neutral and possible to address, even for strangers, e.g. politics, the weather, job, income etc.?
- Common speech acts, e.g. greetings and farewells. Are greetings and farewells always in place or should they be reserved only for some occasions?

6.3 Intercultural variation in perception, understanding and interpretation

Besides cultural variation in the production of communicative behavior, there is also cultural variation in the perception, understanding and interpretation of such behavior. If a male person A does not know that males of group B think that in a normal conversation it is appropriate to stand 10 cm apart (rather than, say, 30 cm), and sometimes touch, their male interlocutors, he might misinterpret what a member of group B does when he steps closer and now and then touches him (A). For an interesting computational model of proximity in conversation, see [4]. In general, all cultural differences in occurring expressive behavior are sensitive to expectations concerning appropriate contents and functions and can therefore be misinterpreted. Since many of the expectations are emotional habits on a low level of awareness and control, they might in many cases, more or less automatically, affect perception and understanding [21]. Thus, a GMIC also needs to have a set of parameters for expectations (e.g. values) and other factors that influence perception, understanding and interpretation.

6.4 Interactive features

Besides parameters for expressive behavior, content, function, and interpretation, other parameters need to be set up to cover variation in interactive features between people with differing cultural backgrounds. Such parameters concern

- Turntaking: How do we signal that speaker change is about to occur? Is it ok to overlap with other speakers? Is it OK to interrupt other speakers? When should interruptions occur? How long should the transition time be from one speaker to the next speaker? Is it OK to do nothing or to be silent for a while in a conversation? What should we do to keep a turn? How do we signal that we don't want the turn, but rather want the other speaker to continue? [22], [15].
- Feedback: How do speakers indicate, display and signal to each other that they can/cannot perceive, understand or accept what their interlocutor is

communicating [19]. Is this done primarily by auditory means (small words like *mhmm*, *m*, *yeah* and *no*) or by visual means (head nods, head shakes, posture shifts etc.) [23], [24]? What emotions and attitudes are primarily used? Is very positive feedback preferred or is there a preference for more cautious feedback? [5].

- Sequencing: What opening, continuing and closing communication sequences are preferred in the culture, e.g. What is the preferred way of answering telephone calls in different activities (opening sequence)? What is the preferred way of ending telephone calls (closing sequence)? When and how should you greet friends and unknown persons when you meet them (opening sequence) [17]?
- Spatial configuration: This includes variation in the size of the distance between the speakers and differences in how speakers orient to each other in different settings (e.g. side by side, face-to-face, 90 degrees etc.)

6.5 Social activity and other kinds of context dependence

Besides the social activity that the communication is part of, there are other contextual features, that can influence communication, e.g. such features may be connected with the deictic features of a language (in English, e.g. words like *I*, *you*, *here*, *now* or tense endings), which in many languages (but not all) are dependent on features of the immediate speech situation. Other factors that might be influential are beliefs, expectations and values that are relevant for several social activities, e.g. ways of showing or not showing respect for persons of another gender, older people or powerful people.

6.6 Impression created in an external observer

Over and above the features of communication introduced above, the behavior of an artificial communicator may also be described according to features introduced by an external evaluator, concerned with establishing whether the behavior of the artificial communicator is believable, responsive and/or interpretable. An evaluation might also be concerned with the quality of what is being simulated, e.g. aspects of cognition, the human body, the environment, emotions, mirroring behavior etc

6.7 A set of parameters for evaluation and suggested functions in an Embodied Communicative Agent

The overview presented above provides us with a number of desirable features in an ECA. They are summarized in table 1 below.

Table 1. Summarizing checklist of communicative features in an ECA

| Features | Specification |
|--|---|
| Activity dependence | goals, roles, artifacts, environment |
| Generic applicability – parameters for cultural adaptation: | |
| Expression: | Eye brow movements Eye movements Arm and hand movements Shoulder movements Intonation in speech Intensity, pitch, duration in speech |
| Content + function | Emotions Attitudes (e.g. politeness) Common speech acts Everyday topics |
| Interactive functions | Turn taking Feedback Sequences Spatial configuration |
| Other types of context dependence | e.g. deixis, beliefs, expectations, values |

7 An example -an evaluation of an artificial communicator used in many cultures – the case of IKEA’s Anna

7.1 Anna in different countries

In order to make our discussion more concrete, we will exemplify it by taking a closer look at an artificial communicator used by a multinational company, IKEA, based in Sweden. We are using IKEA’s Anna as an example of the variation that currently exists in commercial artificial communicators between different countries/cultures. We will also use it to exemplify how the framework introduced above can be used to discuss what could be modified with respect to audiences with different cultural backgrounds.

Anna is an interface to a database of a furniture company. Her main task is, thus, to present web pages with pictures and prices of different types of furniture, but she also provides information about some other aspects of the company. Anna is a fairly simple application, with a neutral-friendly facial expression, some head and posture movements, eye blinks and a very limited set of facial expressions which can be matched to written messages produced by the user or by Anna herself.

The Swedish and “generic” Anna figure is shown in figure 1.



Fig. 1. Anna (Sweden + “generic”)

Whereas Anna’s clothes display the nationality of the company (yellow and blue clothes – colors of the Swedish flag) and indicate selling activity through the outfit of an IKEA sales clerk and with a headset, her skin (fair), hair (red) and eye color (blue) seem to be chosen to show a woman who could come from any European country or North America.

An IKEA web page with an artificial communicator exists in the following parts of the world: *Europe*: Belgium, the Czech Republic, Denmark, Germany, France, Iceland, Italy, Hungary, Netherlands, Norway, Austria, Russia, Poland, Portugal, Switzerland, Slovakia, Finland, Sweden, United Kingdom, *North America*: Canada, United States, *Middle East*: United Arab Emirates/Dubai, *Asia Pacific*: Australia, China, Japan.

The following countries have an IKEA web page without an artificial communicator: Europe: Spain, Greece, Cyprus, Romania, Turkey, *Middle East*: Kuwait, Israel, Saudi Arabia, *Asia Pacific*: Hong Kong, Malaysia, Singapore, Taiwan.

(Data from: IKEA web pages Nov. 2008 and Apr. 2009)

A first question might now be whether IKEA in a particular country chooses to have an artificial communicator like Anna or not. Not all countries have an artificial communicator on their web page. Most European countries, Australia and Japan have an Anna agent and Dubai has a similar agent with darker hair.

The choice of *having or not having an artificial communicator* could clearly be culturally influenced, both with respect to whether it is culturally acceptable or good to have an ECA at all, or specifically a female ECA and with respect to her appearance. We can note that most European countries and Australia have the generic Anna figure and it is IKEA’s official policy to have the same figure. The generic Anna has an appearance, which is typical many women in most of the countries where she appears (Europe and North America). The question of whether to use *a generic or a culturally adapted ECA*, in terms of appearance, is present for all multinational companies.

A next question might be *what an artificial communicator should look like in different cultural contexts*. Here, we can note that three different female agents can be found on the web pages. The Swedish/generic red haired Anna which is most frequent, a blonde “stereotypically Swedish-looking” Anna, which appears on the German and British

web pages (see Figure 2), and a dark haired ECA which appears on the web pages of IKEA in Dubai (see Figure 3).



Fig. 2. Anna in Germany, UK

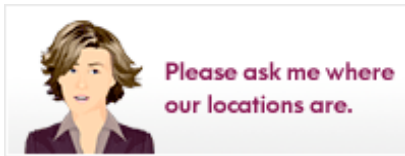


Fig. 3. ECA in Dubai

The blonde Anna emphasizes that IKEA is a Swedish company and links this to the typical image of a Swedish girl, as a sales promoting strategy. The agent in Dubai is perhaps rather an adaptation to the way a modern business-woman might look in Dubai.

If we turn to *the parameters* mentioned above for the behavior of an artificial communicator, we note that Anna moves her head, blinks, uses facial expressions and moves her lips a little when conveying a message. (We will return to this more in detail below.) Her message, however, appears as written text.

Concerning *activity dependence* and roles, the Swedish/generic and the German/British Anna both have the dress of a salesperson and a headset indicating a phone service. She is, thus, adapted to her activity role. The Dubai Anna, on the other hand has a more neutral or formal look, but this might of course be the way an IKEA employee looks in Dubai. She also does not carry a headset. Her picture is accompanied by the question: "Please ask me where our locations are.", so her activity role is clearly a bit different and more restricted than the role Anna has in other places.

Concerning *generic applicability and parameters that can be varied*, there probably are restrictions on what Anna can look like or if an ECA of this type at all is acceptable or even allowed, due to cultural conventions.

If we turn to her conversation, features of feedback, head movements and facial expressions can potentially be very important. Interestingly, for example, the Swedish and Australian Anna do not give equivalent text answers or facial expressions when they are given corresponding input. See Tables 2 and 3.

Table 2. Example of interaction with Swedish Anna

| Input | Answer | Facial expression |
|--|--|--|
| Swedish: Hej (Hi) | Hej hej. Ställ en fråga till mig så ska jag hjälpa dig. Om du söker en produkt går det bra att skriva det åttasiffriga artikelnumret så visar jag mer information om produkten. (Hi, hi. Pose a question to me and I will help you. If you are looking for a product you can type the eight digit number and I will show more information about the product.) | Big smile |
| Swedish: Jag vill köpa en säng (I want to buy a bed) | Här hittar du enkelsängar (Here you find single beds) | Brows lowered, mouth to side, looks doubtful or maybe joking |
| Swedish: Nej (No) | Det var tråkigt (That was sad) | Makes face with her mouth, pulling corners of lips down |

Table 3. Example of interaction with Australian Anna

| Input | Answer | Facial expression |
|---|---|---|
| English (Australia): Hi | Hello there, nice to meet you! | Big smile |
| English (Australia): I want to buy a bed | Please have a look at the beds | Big smile with raised eyebrows (happy look) |
| English (Australia): No | Don't hesitate to speak to me again if you have another IKEA question | Big smile |

Why we find these particular differences is an open question. The facial expressions are chosen among a small number of available alternatives. We can speculate about cultural differences concerning different features of the interaction. Directness would be one variable, where the Swedish Anna goes directly from a short greeting to giving directions about typing in codes etc. The Australian Anna instead gives a longer greeting, possibly indicating a difference in politeness. Both of them smile. To the request for beds, the Swedish Anna gives a surprisingly doubtful facial expression with the direct information about single beds, whereas the Australian Anna provides a

big smile. To the somewhat rude answer from the customer, who says “no”, the Swedish Anna says “That was sad” and makes a quite expressive face with her mouth. The Australian Anna, on the other hand, more or less ignores the “no” and invites a new question with a smile. Do Australians typically display more happiness, more politeness and are they less direct than Swedes? We don’t really know the answer to these questions. The more interesting question is perhaps why there is a difference in interface design. Is the difference based on intuition or on empirical research available to interface designers?

7.2 A checklist of possible and existing features in an ECA like Anna

Table 4. Features of Anna: existing features and suggested improvements, additions

| Features | Specification | IKEA’s Anna | Possible improvement |
|---|--------------------------------------|--|--|
| Activity dependence | goals, roles, artifacts, environment | Very activity specific/ limited | Could be extended within activity. Some everyday topics could be added. |
| Generic applicability - parameters: Expression | Head movements | small movements, no nods, head shakes etc. | Feedback in terms of nods and head shakes could be added. Cultural adaptations could be made of these. |
| | Facial gestures | 3 | Should be extended. |
| | Eye brow movements | In set expressions | Could be made more varied. |
| | Eye movements | - | Could be used more with some recognition of face or gaze, also for directions. |
| | Arm and hand movements | - | Could be added and used for feedback, typical gestures of culture, directions etc. |
| | Shoulder movements | - | Could be added. Maybe not needed for activity or for politeness. |
| | Intonation in speech | in NA(?) text output | |
| | Intensity, pitch, duration in speech | in NA(?) text output | |
| Content + function | Emotions | Has 3 emotions | Should be improved, extended repertoire needed. |
| | Attitudes (e.g. politeness) | Has 3 emotions | Should be improved, extended repertoire needed. |
| | Common speech acts | Has some | Could be extended with respect to some everyday needs. |
| | Everyday | - | Some could be added. |

| | | | |
|---------------------------------|--|--|--|
| Interactive functions | topics | | |
| | Turn taking | Reacts after written message is sent. | Some incremental processing would increase human-like feature and make quicker responses possible. |
| | Feedback | Varied in text + 3 facial expressions | Should be improved considerably, e.g. by added head movements. |
| | Sequences | Only responses to previous request (?) | |
| Other context dependence | e.g. deixis, beliefs, expectations, values | -? | |

7.3 Evaluation of Anna and suggested improvements

Table 4 is an example of how one can go through the checklist given in Table 1, in order to evaluate the features of an artificial communicator and suggest improvements. The checklist can also be used for comparing repertoires of behavior and functions in different artificial communicators.

More advanced agents, like Max, GRETA and others [5], [6], [25], [26] have many of the features mentioned as absent or insufficient (and possible to add or improve) in Anna, in table 4 and that would make her appearance more believable, responsive and interpretable. These agents have a much advanced underlying architecture than most web based agents. Since Anna today is a fairly simple web front-end to IKEA's database, she perhaps does not need as many and advanced functions as the artificial communicators mentioned above have. There are, however, a number of improvements and/or additions that could be made with less advanced methods and that would make her a more pleasant and believable agent. Some feasible and worthwhile changes would be the ones listed below.

- I. Features that would be possible to add without too much added technology making cultural adaptation possible:

The main suggested additions are 1) head movements for feedback (e.g. for *yes* and *no*, positive and negative information and attitudes), 2) some arm and hand movements, which could enhance interpretability by adding redundancy and also could provide deictic information and added expressiveness, 3) an improved and extended repertoire of facial expressions, which can be linked to text output in a more advanced way, and 4) some extended content in terms of frequent everyday topics, which would make her more believable and user friendly.

Motivation for 1-4 above:

1) The addition of head movements, i.e. head nods to go with *yes* and positive information and attitudes and head shakes to go with *no* and negative information and attitudes would make Anna a more pleasant and believable agent.

2) Arm and hand movements are a resource that has not yet been exploited in Anna. They could add to expressiveness and redundancy in information, information structuring. (Shoulder movement is a more debatable feature in an agent with this particular role, since it might be interpreted as impolite, even if it adds expressiveness.)

3) Anna's has three facial expressions, which are holistic composites representing approximately (i) happiness/big smile, (ii) hesitation?, scepticism?, joking attitude?, and (iii) "I'm sorry", "I can't help you", "something is wrong" etc.? They are expressed in the following ways:

(i) Happy: eye brows raised, mouth open with big smile

(ii) Hesitant: eye brows lowered with inner ends lowered, eyes narrowed, mouth closed and drawn to one side

(iii) Sad: Eye brows drawn together with inner ends raised, mouth with lower lip, especially the corners of the mouth lowered, showing teeth ("Making face")

In general, these three facial expressions are too few and too hard to map to the text output to be really helpful, rather than confusing. Facial expressions 2 and 3 are especially hard to interpret. This might be one reason for the choice of the blonde Anna in Germany and the UK, since she does not have expression 2, which seems to be replaced by expression 1 in many cases, making her seem more friendly and polite (this might also be a result of the specific mapping to text). We can also see that the mapping between the facial expression (even of the generic Anna) and the corresponding text messages is not the same in the Anna's of different cultures (see the Swedish and Australian examples above). This could be an attempt at cultural adaptation and it can give this impression, especially in connection with the differences in text responses. The facial expressions would be possible to a) improve and make clearer/less ambiguous, b) extend, making more expressions possible, c) link in better ways to emotions, attitudes and factual information.

In combination with head movements for feedback and possibly some arm and hand movements, improved and added facial expressions would improve Anna's communicative repertoire and "believability". For facial expressions, the studies discussed above can provide information for improvements, e.g. emotions and attitudes could be expressed more efficiently by adding facial expressions, head movements for feedback functions and some arm-hand movements. These features would also add redundancy and thus possibly interpretability to common speech acts.

Concerning content, additions could be made, adding some very common everyday topics and some more topics relevant for an IKEA customer. There are studies with

artificial communicators as front ends to databases in public places, which show typical and frequent questions, requests and attempts at small talk that users initiate and which could be used for identifying a set of topics and typical contributions [26]. The ability to handle at least some of these topics would make Anna more human like and user friendly.

II. Additional suggested features

It would also be fairly simple to add more alternative looks to Anna that might make her look more believable for customers from different cultures. However, this is probably against IKEA's present policy.

III. Features that require more technology and development

Some features that have not been suggested here, since they are more complex and require more research and development than the features mentioned under I and II above are: 1) detection of the user's face, eye gaze or hand movements, which would create a more naturalistic eye gaze and perhaps even make possible some mirroring of behavior, such as nodding or waving, 2) speech output, which would have to be prosodically adapted to the content of written messages and the emotional output of facial expressions.

1) If the user's face, eyes or hands could be detected and followed in space, Anna could be made to direct her eye gaze and provides some response to movements, such as saying *hallo* and *good bye* at the right moment while waving or pointing. This would add to her naturalness and impression of interactive reliability. The ability to use eye gaze and pointing by Anna would also make improved deictic functions possible. However, both these features are demanding with respect to technology research and development.

2) Anna does not have speech output. Speech output would certainly be possible using recordings or a TTS system, whereas speech recognition would be far more demanding and require much more technological development and error management. The prosodic features of the spoken output would probably have to be linked to the written messages and to attitudes and emotions expressed in them in a similar way to what should be done for facial expressions. This would require extra design and resources. The female agent of IKEA in Dubai speaks a pre-recorded text with good intonation, but her speech is not really interactive and her facial expression does not vary.

7.4 Cultural specification of parameters

Given the improvements suggested in I and II above, cultural adaptation could be done with respect to: 1) The text output (adapted to the specific activity in the specific culture), e.g typical sequences and speech acts, choice of word, politeness

etc., 2) Type of feedback words and phrases, showing Contact, Perception, Understanding and Attitudinal reactions (CPUA), 3) What type of response/information to give, where some variables might be the following: formal – informal, long – short, general – specific, direct – indirect, neutral – polite, neutral – expressive, 4) Head movements for feedback, showing CPUA, 5) Facial expressions, CPUA, emotions and attitudes, and 6) Possibly the looks/appearance of the ECA in different cultures.

8 Concluding remarks

In this paper, we have given a first outline of a framework, which attempts to highlight some of the parameters to be taken into account in designing and evaluating a system for multimodal intercultural ICT. We have then exemplified the use of this framework in describing the features of an embodied communicative agent used by an international company in different cultures and suggesting features which could be improved or added.

9 References

1. Peirce, C. S.: *Collected Papers of Charles Sanders Peirce, 1931-1958*, 8 vols. Edited by Hartshorne, C, Weiss, P., Burks. A. Harvard University Press, Cambridge, MA (1931)
2. Allwood, J.: *Capturing Differences between Social Activities in Spoken Language*. In: Kenesei, I., Harnish, R. M. (eds.) *Perspectives in Semantics, Pragmatics and Discourse*, pp. 301--319. John Benjamins, Amsterdam (2001)
3. Kenny, P., Harholt, A., Gratsch, J, Swartout, W., Traum, D, Marsella, S., Piepol, D.: *Building Interactive Virtual Humans for Training Environments*. In: *Proceedings of I/ITSEC (2007)*.
4. Jan, D., Herrera, D., Martinovski, B., Novick, D., Traum, D.: *A computational Model of Culture-specific Conversational Behavior*. In: *Proceedings of Intelligent Virtual Agents Conference*, pp. 45--56 (2007)
5. Kopp, S., Allwood, J., Ahlsén, E., Stocksmeier, T.: *Modeling Embodied Feedback with Virtual Humans*. In: Wachsmuth, I., Knoblich, G. (eds.) *Modeling Communication with Robots and Virtual Humans. LNAI 4930*, pp. 18--37. Springer, Berlin (2008)
6. Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., De Carolis, N.: *GRETA. A Believable Embodied Conversational Agent*. In Stock, O., Zancarano, M. (eds.) *Multimodal Intelligent Information Presentation Kluwer, Dordrecht (2005)*
7. Koda, T.: *Cross-cultural study of avatars' facial expressions and design considerations within Asian countries*. In: Ishida, T., Fussell, S.R., Vossen, P.T.J.M. (eds.) *Intercultural Collaboration I. Lecture Notes in Computer Science*, pp.207--220. Springer-Verlag (2007).
8. Koda, T., Rehm, M., André, E.: *Cross-cultural Evaluations of avatar facial expressions designed by Western and Japanese Designers*. In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) *Intelligent Virtual Agents (IVA 2008)*, LNAI 5208, , pp.245-252, Springer-Verlag (2008).
9. Koda, T., Ishida, T.: *Cross-Cultural Study of Avatar Expression Interpretations*. SAINT 2006, 130-136 (2006)

10. Johansen, S.: Avatars in Global E-Commerce: A Cross-Cultural Analysis of the Effects of Avatars on Online Consumer Behavior in Germany and the U.S. *Journal of Undergraduate Research*, 6(8), 32611; (352) 846-2032 (2006)
11. Kilbourne, J.: *Deadly Persuasion*, New York: Free Press (1999)
12. Hofstede, G.: Cultural Constraints in Personnel Management. *Journal of International Business*, Special Issue 2/98, *Management International Review*, 8-9 (1998)
13. Barber, W., Badre, A.: (1998), "Culturability: the merging of culture and usability. In *Human Factors and the Web*. www.research.att.com/conf/hfweb/proceedings/barber/index.htm, (1998).
14. Chau, P., Cole, M., Massey, A., Montoya-Weiss, M., O'Keefe, R.: Cultural differences in the online behavior of consumers. *Journal of the Association for Computing Machinery*, 45 (10) (2002)
15. Allwood, J.: Are There Swedish Patterns of Communication? In: Tamura, H. (ed.) *Cultural Acceptance of CSCW in Japan & Nordic Countries*. pp. 90–120. Kyoto Institute of Technology, Kyoto (1999)
16. Lustig, M., Koester, J.: *Intercultural Competence: Interpersonal Communication across Cultures*. Longman, New York (2006)
17. Allwood, J., Cerrato, L., Jokinen, K., Paggio, P., Navaretta, C.: The MUMIN Annotation Scheme for Feedback, Turn Management and Sequencing. In: *Proceedings from the Second Nordic conference on Multimodal Communication*. Gothenburg Papers in Theoretical
18. Morris, D.: *Manwatching*. Jonathan Cape, London (1977)
19. Allwood, J.: Bodily Communication – Dimensions of Expression and Content. In Granström, B, House, D., Karlsson, I. (eds.) *Multimodality in Language and Speech Systems*, pp. 7--26. Kluwer Academic Publishers, Dordrecht. (2002)
- Linguistics, 92. University of Gothenburg, Department of Linguistics. (2006)
20. Allwood, J.: *Intercultural Communication*. In: Allwood, J. (ed.) *Papers in Anthropological Linguistics 12*. University of Gothenburg, Department of Linguistics, Göteborg (1985)
21. Hofstede, G.: *Cultures and Organizations: Software of the Mind*. McGraw-Hill, New York (1997)
22. Sacks, H, Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735 (1974)
23. Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9(1), 1-26 (1992)
24. Cassell, J., Thórisson, K.: The power of a nod and a glance. Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 519-538 (1999)
25. Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents. *The Journal Computer Animation and Virtual Worlds* 15(1), 39-52 (2004)
26. Kopp, S.: How Humans Talk to Virtual Humans - Conversations From a Real-World Application. In Fischer, K. (Ed.) *How People Talk to Computers, Robots, and Other Artificial Interaction Partners*, SFB/TR 8 Report No. 010-09/2006, pp. 101-113 (2006)