

Cultural dialects of real and synthetic emotional facial expressions

Zsófia Ruttkay

Received: 30 September 2008 / Accepted: 27 May 2009 / Published online: 1 July 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract In this article we discuss the aspects of designing facial expressions for virtual humans (VHs) with a specific culture. First we explore the notion of cultures and its relevance for applications with a VH. Then we give a general scheme of designing emotional facial expressions, and identify the stages where a human is involved, either as a real person with some specific role, or as a VH displaying facial expressions. We discuss how the display and the emotional meaning of facial expressions may be measured in objective ways, and how the culture of displayers and the judges may influence the process of analyzing human facial expressions and evaluating synthesized ones. We review psychological experiments on cross-cultural perception of emotional facial expressions. By identifying the culturally critical issues of data collection and interpretation with both real and VHs, we aim at providing a methodological reference and inspiration for further research.

1 Introduction

1.1 Toward culturally adaptive virtual humans

One of the major motivations for developing virtual human (VHs) is their potential as the most natural and easy to use interfaces to computer services, making these accessible for a broad range of users. In our connected and globalized world, the majority of public application would or could

attract a culturally diverse user group. For instance, a holiday booking assistant, or a coach to help the user to stop smoking has to deal with very different clients. In everyday life the success of people in such roles very much depends on how well they ‘speak the language, or find the words’ of their client. By speaking the language we refer to all aspects of communication beyond the literal language usage (which, of course, should be familiar to the client). They span from finding out the values and goals of the client to frame the task at hand (what to say), the subtleties of language usage (colloquial or not), to accommodating the control of conversation, the adjustment of speech tempo and maybe dialect, and the facial and bodily expressions. On the other hand, people, even if trained in communication, are limited in their degree of adaptation, as constrained by their bodily and facial features, their personality and their cognitive resources (such as the languages they speak).

When developing a VH for an application, we could, in principle, accommodate a design not only to match, but also to outperform real people in terms of adaptation capabilities. All that is needed:

- assessment of the culture of the user;
- ‘instantiation’ a VH with the bodily, cognitive, and communicative capabilities best for the user (and of course, the task) at hand;
- letting this VH appear for the user and carry out the tasks further (including possibly learning about the user and further adaptation to him/her).

This scenario is far from feasible at the moment. On the one hand, necessary technological components—image, speech and language processing, handling of huge common sense knowledge bases—are not yet powerful enough. On the other hand, we are lacking the design principles and

Z. Ruttkay (✉)
Department of Computer Science, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: zsofi@cs.utwente.nl

evaluation methodology to find out what we would go for, had we these technological components at our service. It is evident that to answer such a question (further) research from social and behavioral psychology, cultural anthropology, and psycholinguistics is needed. At the same time, the perspective of seeing a VH in a cultural context casts new light on some of the existing results, and makes it possible to set a framework to systematically investigate and design for cultural differences.

1.2 The cultures to be considered

We interpret culture from a pragmatic point of view as a set of characteristics which form a ‘common denominator’ among groups of people, including both mental and communicative characteristics, as values in life and multimodal language usage. Hofstede’s (2001) seminal work aimed at characterizing *societies and nations*. Racial features for embodiment and the accommodation of cultural (multimodal) language usage are a starting point to design cultural VHs. A statistical characterization of a large society is inevitably too stereotypical—one is not dealing with an average American or Italian, but an American Jewish professor from NY City, or a farmer from Texas, or a Sicilian fishermen, or an elderly women from Naples. *Subcultures* can be identified based on religion, education, profession and social status, or origin from a specific region of a country. Though *age and gender* seem to be clear-cut biological parameters, in societies these imply differences in view on life, mental and communicative behavior, and in different cultures are important factors of communicational protocols, indicated also by the masculinity–femininity dimension of Hofstede’s definition of national culture. A culture may be *trans-national* not related to ethnic identity (Hannerz 1992). We talk about the ‘youth culture of today’ or the Western culture.

Culture is manifested in all levels of processes guiding social interaction (Mesquita et al. 1997), and inversely, people assign a cultural background to (virtual) humans based on the look, the language usage, the views manifested in conversation. The necessity of cultural adaptivity for embodied agents has been identified early (O’Neill-Brown 1997). Nass et al. (2000) proved that ethnic in-group identity of the VH, manifested only in the look, influenced the judgment of the VH. See Payr and Trappl (2004) for in-depth discussion of the cultural differences in behavior rules, and a few case-studies of agent systems for multicultural applications and general considerations for designing such agents and ongoing projects are aiming at testing the perception of culturally specific VHs (Maldonado and Moares 2002; Iacobelli and Cassell 2007; Koda 2004; Rehm et al. 2007).

1.3 Focus on emotional facial expressions

In this article we concentrate on the issue of emotional facial expressions and culture. The choice of our focus is twofold. On the one hand, one cannot design a VH without *some* emotional facial expressions. Which expressions should be considered, from a semantic point of view, and how should the face and its repertoire be designed, if the VH is to be used in a specific culture? What methodology should be used to gather the necessary, culturally representative samples? How should we evaluate facial expressions on a VH? How should individual facial behavior within a culture be designed, to avoid ‘cultural stereotypes’?

Another motivation for concentrating on facial expressions is that there has been a substantial body of literature in psychology on the display and perception of facial expressions, providing theories for computational models. Also, there are some emotional facial expression databases for references (http://vasc.ri.cmu.edu/idb/html/face/facial_expression/; Ekman and Friesen 1976; <http://kasrl.org/jaffe.html>). Facial expression recognition technology is reaching the stage of real-time recognition of spontaneous expressions in every-day environments (Bartlett et al. 2006). Hence there are resources and tools to gather culturally specific data on facial expressions—the question is how to use them.

In the rest of the article, we will discuss the implications of culture on the design of facial expressions for VHs. In the next section, we provide a formal framework to flesh out the steps involved in designing facial expressions for VHs and identify the culturally sensitive aspects, in general. We clarify the concepts of the facial display space and the emotions space, and discuss similarity measures and mappings on these spaces as references for comparing cultural differences. Further on, we investigate, by quoting empirical psychological studies and VH experiments, how cultural factors play a role in display and perception of emotional facial expressions. Finally, we sum up the recommendations for culturally specific facial expression design and raise some general questions concerning the usage of culturally adaptive VHs.

2 Culturally sensitive steps in the process of generating facial expressions

2.1 Stages of designing facial expression

Adapting the general analysis—design—evaluation cycle also for VH design (Ruttkey and Pelachaud 2004), endowing VHs with facial expressions requires the accomplishment of the following steps (see Fig. 1):

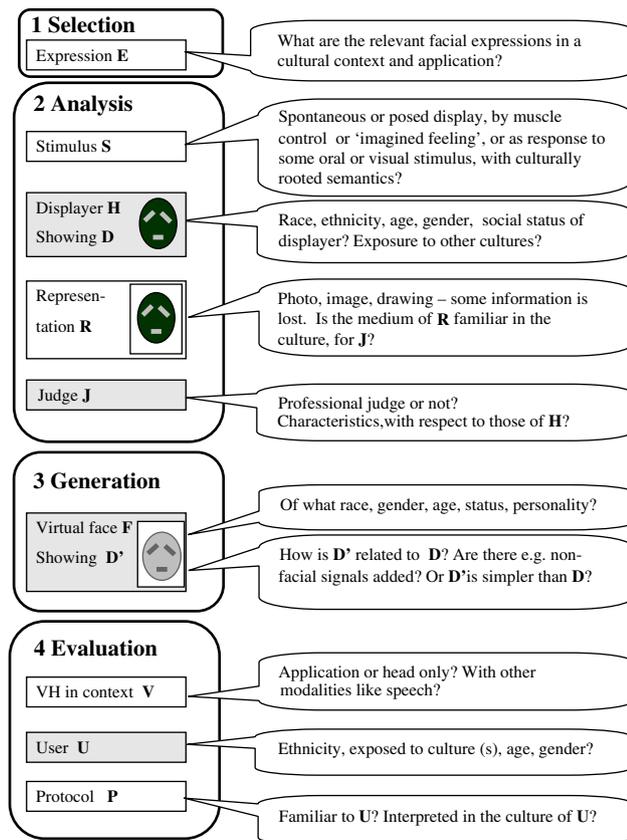


Fig. 1 Identifying culture-related questions in stages of designing facial expressions for VHs. The (real or virtual) human actors are indicated in *gray boxes*, data and representations in *white boxes*

1. *Select* the expression **E** of interest (e.g., happiness).
2. *Analyze* how this expression is produced in real life, by:
 - a. gathering facial display samples **D** performed by **H** humans, using **S** stimuli;
 - b. creating an **R** representation (such as video or photo) of the displayed samples;
 - c. peer-reviewing them by human judges **J**, assuring that **D** (shown in some representation **R**) is perceived as display of the emotion **E** of interest.
3. *Generate* the facial expression, considering:
 - a. a virtual facial model **F**;
 - b. showing the facial expression **D'** (meant to be the synthesis of **D**).
4. *Evaluate* how the generated expression is perceived, by:
 - a. creating the representation **V** (for example, showing the VH in an application context such as talking head telling a story or a tutor guiding a study, or showing the facial expressions only);

- b. gathering the interpretation **E'** by potential users **U**, by using some semantic evaluation protocol **P**.

Ideally, one would like to have **E'** be identical to **E**, meaning that the generated facial display on the virtual face conveys to the user the expression as it is intended to. Often this is not the case. Here we set out to discuss only those possible causes of mismatch which have to do with some cultural factors. Obviously, the people involved as *actors* (in the role of displayer, judge, or user) introduce per se a cultural component. However, the *data* and its representation used in the process may also introduce, in an indirect way, cultural biases. In Fig. 1 we visualize the steps of the process, highlight (virtual) humans involved and the basic data used, and at each step raise questions to identify potential causes of a mismatch between intended and finally perceived facial expressions. In the discussion further, we will refer to the (virtual) humans and other factors by the letters also shown in Fig. 1.

Some clarification for the above protocol: in the analysis stage (step 2), we did not differentiate between 'psychological study' and 'empirical data analysis', as in the rest of the article we will look at the original empirical analysis performed by psychologists, used to propose or justify a theory. Note also that the general protocol and related questions apply for any facial expression, not only for emotions.

In the analysis step, we restrict ourselves to data gathered from real people (discarding, e.g., artistic rendering of facial expressions in animations or in paintings). In the synthesis stage, on the contrary, non-realism and additional features to enhance the facial expression may be used, because of the very nature of the virtuality of the humans.

It is evident that cultural factors may have an influence whenever a real or VH is involved, that is: **H + D**, **J**, **F + D'**, **U**. Actually, the question "What does the face reveal?" is to be answered by referring to the perceiver too, as remarked by the advocates of the *relational model* of facial expression recognition (Elfenbein and Ambady 2003).

2.2 Mapping from facial displays to emotions

To answer the question if different cultures use similar facial displays for similar emotions, we need notions of similarity in the two separate spaces: **D** the *space of facial displays*, and **E** the *space of emotions* which may be attributed to them (see Fig. 2).

3 Similarity of facial displays

Numerical coding systems such as FACS (Ekman and Friesen 1978) or MPEG-4 (Pandzic and Forchheimer 2002) allow the objective and face-independent coding of facial

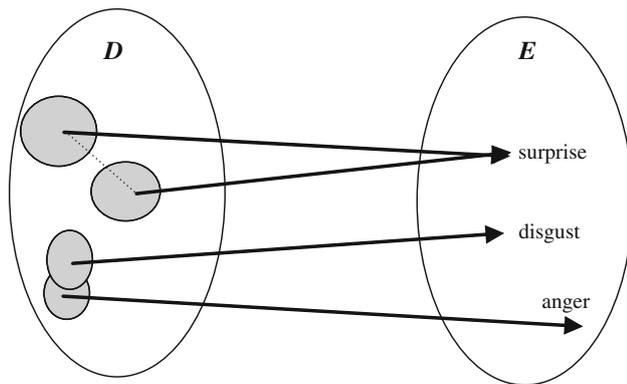


Fig. 2 Mapping from space of facial displays D to the space of emotions E . Interpolation between two ways of displaying surprise and proximity of display of disgust and anger are shown in D

display, resulting in a high-dimensional vector $v = v(D)$. The dimensionality of the vector corresponds to the number of parameters used to code expressions. For instance, when using all the MPEG-4 Facial Action Parameters (FAPs) to describe facial displays, the space D is 68 dimensional, and points in this space correspond to vectors of FAP values describing facial displays. Of course, only a subset of the entire D corresponds to displays which can occur on faces (this is also reflected by the limits for the individual parameters), and only some of these are expressive, meaningful. Note that even when using all the 68 parameters, compared to reality, substantial information on the visual appearance of a real facial expression is discarded, such as tears in the eye, blushing, and humidity of the face. These factors may very well contribute to the judgment of expression on real faces. This is suggested by the study showing systematically lower expression recognition accuracy on a state-of-the-art, textured talking head, compared to recognition rate achieved on photos from databases showing real humans with identical facial displays (Kätysri et al. 2003).

To compare (unlabeled) facial displays, some *similarity measure* is to be used. One is free to choose from the arsenal of measures of numerical spaces, see for example (<http://www.dcs.shef.ac.uk/%7Eesam/simmetrics.html>) for a summary. These measures, according to their mathematical definition, all ensure relational symmetry in the similarity (A is as similar to B as B is similar to A) and the triangle inequality: A and C are at least as similar as the similarity of A and B plus the similarity of B and C . While there is supportive data on the relational symmetry of perceived similarity of faces (Niewiadomski and Pelachaud 2007), we do not know if the triangle inequality has been tested, and thus if it should be imposed. In the definition of some measure one is free to choose some parameters. For example, the weighted distance assigns weights to the different dimensions. When applying it to the display space

D , one may choose, for instance, bigger weights to eyebrow deformation parameters than to mouth deformation parameters.

Which measure of similarity is the best? This question suggests that there is some oracle who can judge the ultimate similarity of facial expressions, and we must find the measure which approximates this ultimate judgment. We do not have such a single oracle but we do have people's opinions. So we may be able to show, experimentally, that one measure coincides better with the judgment of a single person, or of a group of people acting as judges in experiments. We know of a single work addressing this issue directly (Niewiadomski and Pelachaud 2007). The authors define a measure for D by 'fuzzifying' each expression by using symmetrical trapezoid functions for each parameter, and then using a measure in the space of the fuzzy sets as the indication of similarity of facial expressions. All parameters (and facial regions) contribute equally to the similarity measure. The authors did not consider other possible measures (e.g., the non-fuzzy version of the fuzzy one) to justify their choice. They showed that their chosen measure correlated well with human judgments of similarity of expressions. They noticed, at the same time, that their measure was not uniformly coinciding with human perceptions, in all regions of similarity and for all pairs of facial displays. Judges were collected via the web, and the potential influence of the (cultural) characteristics of the judges was not discussed.

A systematic investigation on comparing different measures—similarly to how different measures are considered in computer facial recognition techniques—may help to answer questions such as: Do distinct facial regions contribute differently to perceiving differences? What is the influence of the absolute and of the relative value and of the possible magnitude of the different parameters? How small the changes are that people notice as different? How is the sensitivity related to what they see around themselves in everyday life? For example, does asymmetry in the facial display—such as an asymmetrical eyebrow raise—enhances the perceived effect of a single parameter? Is there a difference in judging the similarity of 'meaningful' and 'meaningless' facial expressions?

Investigation of such questions can be very useful for both psychologists giving further insight to how people look at facial features and for the designers of expressive VHs. Our own earlier studies show that asymmetric facial expressions designed by a trained graphical artist were much better identified by subjects, than the usual symmetrical variants captured from real faces (Hendrix and Ruttkay 2000). A possible explanation for this phenomenon may be that the asymmetric shapes trigger more attention on the lowest level of perception of faces. Another outcome of such experiments would identify the

characteristics of people who seem to use ‘similar measures’. Women are known to be better in interpreting facial expressions (Montagne et al. 2005)—maybe they already notice differences on a smaller scale? Some cultures are ‘more gazing’ than others—does this imply that these cultures use different measures of similarity of faces? For example, do non-gazing cultures notice more changes on the non-eye region of the face? How the ‘facial dynamism’ of the culture of the judge does influence his sensitivity to differences?

Finally, we must notice that already in this stage of evaluating differences of non-labeled facial displays, we have to take into account the *very face* on which the expressions are shown (**H** and **F**). A judge (**J** or **U**) is influenced by the face too, not only by the expression it displays. He/she may be more or less motivated, depending on the gender of the face, the in-group character (age, ethnicity). The familiarity with the facial physiognomy may result in an own-race bias in facial perception. For more details see (Hirose 2006).

3.1 Similarity of emotional facial expressions

Ultimately, one is interested in what meaning—particularly, what emotion—a facial display evokes. Different displays may convey identical meaning. For example, surprise may be displayed with eyebrow raise alone, with open mouth alone, or by both, with different intensity of the eyebrow and mouth features. Hence different points of the facial display space may convey *identical* emotions, or different *intensities* of emotions, or *similar* emotions (see Fig. 2). While we all agree that disappointment is an emotion similar to sadness but very different from happiness, it is not straightforward how to measure the similarity of emotions. The facial display can be coded in terms of absolute values according to accepted protocols which can even be automated, but there are no such accepted coding mechanisms for emotions. One may use a *continuous appraisal model*, where points in a 2 or 3 dimensional space correspond to emotional states (Russell 1980; Ruttkay et al. 2003), or use a *categorical model*, where emotions are identified by discrete labels (Ekman and Friesen 1975). Biological signals—such as skin conductivity, heart-rate, or different measures of brain activity (Güntekina and Basar 2007)—may be tapped to indicate arousal, but they are not rich enough or universal enough to be able to derive different emotional states precisely from them (Nakasone et al. 2005). The definition and testing of similarity of emotions is much more problematic than the testing of similarity of visual displays. Some facial displays, such as the general displays of the six basic expressions described by Ekman (1992) and Ekman et al. 1987 are ‘hard-mapped’ to the emotional expression space.

We know the emotional meaning of these points in the display space, allowing alternative displays of the same emotion. The facial signal-emotion mapping is thus a function, of which only a few values, typically, those of the six basic expressions, are known. The following questions arise:

Is the display-emotion mapping continuous in the sense that *similar displays correspond to similar meanings*?

Does distance from neutral expression in the display space correspond to *intensity of expressions* in the expression space?

If \mathbf{v}' and \mathbf{v}'' —two points in the display space \mathbf{D} —both express the same emotion, how about the *linear combination* of them (corresponding to a line connecting these two points in \mathbf{D})?

In the other direction, what can we say about the *mixture of expressions*, such as pleasant surprise? Some researchers have been using different principles and rules to create displays of mixed expressions from the display of the single expression on synthetic faces, such as assigning different facial regions for the partial display of the positive–negative expressions (Martin et al. 2006), or adding up and normalizing the parameters values of the distinct expressions (Ruttkay et al. 2003).

More generally, what *regions of the display space are perceived as a certain emotional expression*? We believe that this mapping is rather complex and should reflect correlation and constraints between feasible parameters.

What are the *individual differences*? How could we design individual repertoires in terms of modifying this mapping? E.g., by making the display region assigned to different emotions smaller/bigger or replacing entirely some regions? Where are the *exaggerated expressions* in \mathbf{D} ?

In establishing the display-meaning mapping, and particularly, in modeling emotions in a fuzzy way, we need to be clear about two aspects:

- how *intense* is the emotion perceived;
- how *unanimously* is the emotion perceived.

There is a difference between subjects agreeing that a facial display shows ‘a little surprise’ on a scale factor 4, and 0.25 of the subjects thinking that a facial display shows surprise, 0.75 judging it as neutral.

The latter example raises the issue of whether disagreement on judgment could be considered as a fuzzy measure of displaying some meaning.

How to interpret the fact that certain emotions are ‘easier’ to recognize than others? For instance, is ‘smile’ less fuzzy than ‘disgust’? Facial display of disgust and anger are often mistaken—does this have to do with similarity, in terms of facial display, of the two expressions, while the smile expression is far from the other five in the facial display space? Our own investigation of \mathbf{D} suggested

small distance between the mistaken negative facial displays (Hendrix and Ruttkay 2000), and a machine vision system has produced similar error patterns in recognizing facial expressions (Dailey et al. 2002), so it may be the case that humans may use a similar measure to compare facial displays.

The judgment of emotional meaning of facial displays raises the same concerns as the judgment of similarity of unlabeled facial expressions. Moreover, the elicitation of emotional labels is more prone to cultural aspects, than the notion of similarity, discussed in more detail in the next section.

4 Empirical studies on cultures and emotional facial expressions

4.1 Cultural dialects of displaying universal emotions

Ekman and his colleagues have collected a huge body of experimental data to support the universality of the six basic expressions (Ekman 1992; Ekman and Friesen 1975; Ekman et al. 1987). Critics of their methodology (using forced choice test) and results (higher error rates with non-Western cultures) proposed a continuous model as opposed to discrete categories (Haidt and Keltner 1999; Russell 1980, 1994, 2003; Schiano et al. 2004). If we realize how culture-dependent the factors are in the display and recognition process outlined above, the two, seemingly antagonistic theories can be bridged. Hess and co-workers (Elfenbein et al. 2007) have coined the term *cultural dialects of facial expressions*: the cultural dialect, unlike a personal idiosyncratic variant, is a well identifiable specific usage of some signal. Similarly to language dialects with specific words used with specific meanings as well as systematic deviations in pronunciation or grammar usage, one can find culture-specific emblems for certain emotional display as well as variants in display rules (e.g., being less articulated in facial displays, or not showing negative emotions). And just as it is with language dialects, one can get used to a facial dialect and understand it better, even accommodate it—this is becoming a necessity in our multicultural life. The importance in language dialect is well proven in applications with English synthetic speech, where a ‘Chinese pronunciation dialect’ is offered for Chinese users as opposed to insisting to the normative Oxbridge pronunciation.

To bring these dialects to light, and to show their role in choosing H, J, F, and U, a big body of work on cultural differences in displaying and/or interpreting facial expressions compares Asian and European or American subjects.

Kito and Lee (2004) found that the British were inferior to the Japanese in interpreting interpersonal relationships

based on facial display of Japanese people in photos. Elfenbein and Ambady (2003) showed that the recognition of the six basic expressions was always above chance level when judged by people of different cultures, but the recognition rate was higher when the (ethnic or regional) culture of the persons displaying and recognizing the expression were identical. This racial bias was less if the perceiver had extensive contacts with other cultural examples of facial expressions. The same authors carried out in-depth research to trace how ‘cultural exposure’, and familiarity from everyday life, influences facial emotion recognition (Elfenbein and Ambady 2003). They looked at accuracy and speed of recognizing Chinese and American facial expressions by groups of Chinese ‘exposed’ to these two cultures differently based on how long they had been living in the US. Photos of displays of the six basic expressions, performed by Americans and Chinese in their country of origin, were used for comparison. Interestingly, Chinese after spending 2.4 years in the US were better at judging the emotional expressions of American faces than of Chinese faces. Similar effect of cultural exposure was found when looking at Tibetans living in China and Africans living in the US. As to comparing general characteristics of facial expression recognition, participants in China were less accurate and slower than participants in the US. This study underlines the inevitable impact of *learning* on interpreting culturally different facial expressions. The authors even suggest that this learning may be very much motivated when the nonverbal signals are the only way to judge others’ emotions as their language is not understood. However, differences in the experimental setting might have influenced the outcome. The two databases of photographs were created as a ‘facial recognition benchmark’ and as a ‘socially appropriate expression’ collection in the US and China, resulting in differences in the intensity of expressions. Furthermore, the lower socioeconomic background of the Chinese subjects, students coming from a less reputed university than the American subjects, probably also played a role. It is also noted that the different display rules make Chinese people less attuned to facial expressions, altogether. Finally, as the ethnical group of the posers was evident from the photos, it could have biased the judgment of the expressions in different ways. Judges may be more motivated to interpret facial expressions of people they identify with; they may use stereotypical or ‘reasoned’ judgments for expressions on the faces of people from other cultures. Finally, the English language of the experiment may have led to a bias towards the judgment of American facial expressions. The latter effect was shown for Indian participants when evaluating facial expressions in English and in Hindi (Matsumoto and Assar 1992).

Same bias toward own race was reported within subjects born in the US but with different racial background

(Matsumoto 1993). The in-group decoder bias was present even when the ‘culture’ was being a basketball player (Thibault et al. 2006): basketball players were more accurate in decoding facial expressions from faces when they were told that they were looking at a photograph of a basketball player than when the face belonged to a non-player. The labels were assigned randomly to faces (of non-basketball players) displaying emotional expressions.

In Bartneck et al. (2004) two kinds of culturally neutral, cartoon-like simple faces displayed a range of expressions, which were judged by Japanese and Dutch subjects. It was clear that the difference in the facial design influenced the judgment of the identical and dynamical expressions. The cultural difference was to be noticed as interpretation differences due to display rules or due to different meaning of some symbolic gestures. Moreover, Japanese women were more positive about and sensitive to the displayed expressions.

In Abelin (2004) it was shown that static facial expressions shown on cartoon-like faces improved how Swedish subjects could recognize the emotional content of Spanish speech. The facial expressions were also used as stimuli to create the emotional intonation to avoid linguistic and categorization problems.

4.2 Protocols to identify displayed emotions

Different facial expressions are usually described and identified by labels, such as angry, frustrated, upset, sad, or disappointed. A commonly used methodology (**P**) is to force the judge or the user to choose a *single label* per expression from an exclusive set, as opposed to offering a set of labels for ‘similar’ feelings, or leaving it to the judge to describe the expression freely, and afterward the experimenter is to analyze the free description and conclude about a category.

It is not certain whether the English labels have an equivalent in another language. It has been argued that English has the biggest emotional vocabulary, and that this richness of the language may be the cause of the superior capability of Americans in understanding emotions also form facial displays (Matsumoto and Assar 1992).

Alternatives to labeling have been used, such as asking ‘what has happened to the person showing the facial expression’ and analyzing the story afterward. Note that here too the analyzer must be closely familiar with the culture, to interpret the story in the cultural context of the person reacting, and not of his own. Another alternative to bridge the language gap is to use simplified drawings of facial expressions as triggers for display (**S**), and as labels to identify them (**P**) (Abelin 2004). One may wonder, though, how universal such cartoon facial displays are. Moreover, by giving visual cues, the display and the

interpretation may be biased by trying to match the real faces to the cartoon drawings eliminating the emotional level entirely.

As to *triggering the facial display* by some stimulus **S**, there is a difference if spontaneous expressions are recorded, or if they are posed by non-actors or actors, or by trained FACS coders who can master the coordinated, conscious control of the different facial muscles. The difference in the triggering mechanism results in different quality of display of the same emotions in culturally different databases (Kätsyri et al. 2003). The culturally specific display rules (Matsumoto 1990) can add a layer between perceiving a facial expression and associating it with an emotion.

4.3 Age and gender of raters

Several studies underline that emotion recognition becomes faster with age without losing quality in accuracy (Kestenbaum and Nelson 1992). One may wonder how persistent this finding is in cultures where youth is overwhelmed with visual stimuli, and has more experience with coming into direct contact with other cultures, by traveling.

Another common finding is the superiority of women in decoding facial expressions. Hence the age and gender distribution of different sets of subjects judging facial displays (**J** or **U**) should be similar.

4.4 The face displaying expressions

Cultural dialects of facial expressions are tested with real faces which bear the characteristics of a given race. It is hardly possible to have an American being able to show an expression in a natural way, as it is performed in Japan. However, using virtual characters, it would be possible to show coded facial expression (such as a smile performed by an American) on faces of different ethnicity. In Hirose (2006) morphs of Japanese and British faces were used to create in-between racial variants to test the race effect on different facial perception tasks.

4.5 Experimental protocols

As systematic study of facial expression has been carried out by researchers in Western institutions, the protocols (**R** and **P**) like rating photos or using computer evaluation settings are taken as normal. With people from cultures or social groups who are not familiar with looking at photos, or participating in scientific research experiments, the experimental setting may be an extra threshold. Also, the presence of an experimenter, especially from another culture, may influence the reactions of the subjects.

5 Discussion

We took a careful look at the steps involved in designing emotional facial expressions for virtual characters. We showed, by referring to studies from social psychology and experiments with VHs, how the cultural aspects must be taken into account for the people involved in displaying (**H** and **F**) and rating (**J** and **U**) expressions, as well as the culture of the experimenter interpreting the results. Moreover, we showed that one would like to have an ‘identical setting’ in triggering the facial expressions (**S**), and running the evaluation experiments (**V** and **P**), which may introduce a Western bias to cultures.

We pinpointed the essential (and vulnerable) aspects of experiments to gather data to design and test culturally specific virtual agents. We hope that this knowledge will help forthcoming research converge as to methodology and interpretation of results gained in different settings, and will serve as a basis to develop VHs with faithful cultural communicational traits, including display of emotions.

The VH technology may offer new in some sense culturally neutral means to learn more about the culture-specific aspects of facial expressions. One option is to retarget facial expressions from one culture to faces of another, thus separating the bias of the racial characteristic of the face.

As for designing VHs for a multicultural public, different strategies may be chosen:

1. For each user, create on the fly the culturally matching VH.
2. Use a single design with an ethnical identity (e.g., Caucasian), but adopt the communication rules of the culture of the user.
3. Use some non-realistic, ‘above culture’ characters and communication capabilities, possibly enhanced with non-realistic symbols.

We are not aware of any working system of the first type, but of case studies showing the benefits of culturally matching virtual conversant in, for example, learning environments (Iacobelli and Cassell 2007). The second strategy is to be chosen in situations when a ‘foreigner’ is to communicate with locals, and his different cultural background is relevant (e.g., he is representative of a peace keeping force (Traum et al. 2005), or a salesperson from a foreign company). The last scenario assumes the very exciting possibility that non-realistic virtual characters may introduce an own culture of communication, which will be learnt by the youth who will be exposed to communicating with such agents, similar to the usage of emoticons in e-mails.

In the above scenarios, a basic decision is whether we want to design for the ‘best match’ for a user, resulting in services where the user will succeed with minimal mental effort—but would this not lead to an impoverishment of the

experience, not stimulating the user to stretch his own mental and communicative capabilities? How should we design for ‘just the right amount’ of mismatch? These general questions ask for further research.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abelin A (2004) Cross-cultural multimodal interpretation of emotional expressions—an experimental study of Spanish and Swedish. In: Proceedings of speech prosody, Nara, Japan, 2004. <http://www.isca-speech.org/>
- Bartlett MS, Littlewort GC, Frank MG, Lainscsek C, Fasel I, Movellan JR (2006) Automatic recognition of facial actions in spontaneous expressions. *J Multimedia* 1(6):22–35
- Bartneck C, Takahashi T, Katagiri Y (2004) Cross-cultural study of expressive avatars. In: Proceedings of social intelligence design 2004, Enschede, pp 21–27
- Cohn-Kanade AU-Coded Facial Expression Database http://vasc.ri.cmu.edu/idb/html/face/facial_expression/
- Dailey M, Cottrell G, Padgett C, Adolphs R (2002) EMPATH: a neural network that categorizes facial expressions. *J Cogn Neurosci* 14(8):1158–1173
- Ekman P (1992) An argument for basic emotions. *Cogn Emot* 6:169–200
- Ekman P, Friesen WV (1975) *Unmasking the face*. Prentice Hall, Englewood Cliffs
- Ekman P, Friesen WV (1976) *Pictures of facial affect*. Department of Psychology, San Francisco State University, San Francisco
- Ekman P, Friesen WV (1978) *The facial action coding system*. Consulting Psychologists Press, San Francisco
- Ekman P, Friesen WV, O’Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K (1987) Universals and cultural differences in the judgments of facial expressions of emotion. *J Pers Soc Psychol* 53:712–717
- Elfenbein H, Ambady N (2003a) Cultural similarity’s consequences, a distance perspective on cross-cultural differences in emotion recognition. *J Cross-Cult Psychol* 34(1):92–110
- Elfenbein H, Ambady N (2003b) When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J Pers Soc Psychol* 85(2):276–290
- Elfenbein H, Ambady N (2003c) Universals and cultural differences in understanding emotions. *Curr Dir Psychol Sci* 12(5):159–164
- Elfenbein H, Beaupré M, Leveque M, Hess U (2007) Toward a dialect theory: cultural differences in expressing and recognizing facial expressions. *Emotion* 7:131–146
- Güntekin B, Basar E (2007) Emotional face expressions are differentiated with brain oscillations. *Int J Psychophysiol* 64(1):91–100
- Haidt J, Keltner D (1999) Culture and emotion: multiple methods find new faces and a gradient of recognition. *Cogn Emot* 13:225–266
- Hannerz U (1992) *Cultural complexity: studies in the social organization of meaning*. Columbia University Press, New York
- Hendrix J, Ruttkay Zs (2000) *Exploring the space of emotional faces of subjects without acting experience*, CWI Report INS-R0013, Amsterdam
- Hirose Y (2006) *The effect of facial expression and identity information on the processing of own and other race faces*, PhD Dissertation, University of Sterling

- Hofstede G (2001) Culture's consequences, comparing values, behaviors, institutions, and organizations across nations. Sage Publications, Thousand Oaks
- Iacobelli F, Cassell J (2007) Ethnic identity and engagement in embodied conversational agents. In: Proceedings of intelligent virtual agents (IVA), pp 57–63
- JAFFE: Database of Japanese Female Facial Expression <http://kasrl.org/jaffe.html>
- Kätsyri J, Klucharev V, Frydrych M, Sams M (2003) Identification of synthetic and natural emotional facial expressions. In: Proceedings of the AVSP'2003, pp 239–244
- Kestenbaum R, Nelson CA (1992) Neural and behavioral correlates of emotion recognition in children and adults. *J Exp Child Psychol* 54:1–18
- Kito T, Lee B (2004) Interpersonal perception in Japanese and British observers. *Perception* 33(8):957–974
- Koda T (2004) Interpretation of expressive characters in an intercultural communication. In: Negoita MG, Howlett R, Jain LC (eds) Knowledge-based intelligent information and engineering systems (8th international conference, KES2004). Lecture notes in artificial intelligence, vol 3214, part II, pp 862–868
- Maldonado H, Moares M (2002) Designing for diversity: multi-cultural characters for a multi-cultural world. In: Proceedings of the of IMAGINE 2002, Monte Carlo
- Martin C, Niewiadomski R, Devillers L, Buisine S, Pelachaud C (2006) Multimodal complex emotions: gesture expressivity and blended facial expressions. *Int J Hum Robot* 3(3):269–291
- Matsumoto D (1990) Cultural similarities and differences in display rules. *Motiv Emot* 14(3):195–214
- Matsumoto D (1993) Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motiv Emot* 17(2):107–123
- Matsumoto D, Assar M (1992) The effects of language on judgments of universal facial expressions of emotion. *J Nonverbal Behav* 16:85–99
- Mesquita B, Frijda NH, Scherer KR (1997) Culture and emotion. In: Berry JW (ed) Handbook of cross-cultural psychology. Basic processes and human development, vol 2. Allyn & Bacon, Boston, pp 254–297
- Montagne B, Kessels R, Frigerio E, de Haan E, Perrett D (2005) Sex differences in the perception of affective facial expressions: do men really lack emotional sensitivity? *Cogn-Process* 6:136–141
- Nakasone A, Prendinger H, Ishizuka M (2005) Emotion recognition from electromyography and skin conductance. In: Proceedings of the 5th international workshop on biosignal interpretation, Tokyo, Japan, pp 219–222
- Nass C, Isbister K, Lee E (2000) Truth is beauty: researching embodied conversational agents. In: Cassell J, Sullivan J, Prevost S, Churchill E (eds) Embodied conversational agents. MIT Press, Cambridge
- Niewiadomski R, Pelachaud C (2007) Fuzzy similarity of facial expressions of embodied agents. In: Proceedings of the IVA07, pp 86–98
- O'Neill-Brown P (1997) Setting the stage for the culturally adaptive agent. In: AAAI fall symposium on socially intelligent agents, pp 93–97
- Pandzic I, Forchheimer R (2002) MPEG-4 facial animation: the standard, implementation and applications. Wiley
- Payr S, Trappl R (2004) Agent culture. Lawrence Erlbaum Associates, London
- Rehm M, André E, Bee N, Endrass B, Wissner M, Nakano Y, Nishida T, Huang H (2007) The CUBE-G approach—coaching culture-specific nonverbal behavior by virtual agents. In: Proceedings of the 38th ISAGA, Nijmegen
- Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39:1161–1178
- Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull* 115:102–141
- Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110:145–172
- Ruttkay Zs, Pelachaud C (eds) (2004) From brows till trust: evaluating embodied conversational agents. Kluwer
- Ruttkay Zs, Noot H, Ten Hagen P (2003) Emotion disc and emotion squares: tools to explore the facial expression space. *Comput Graph Forum* 22(1):49–53
- Schiano D, Ehrlich S, Sheridan K (2004) Categorical imperative NOT: facial affect is perceived continuously. CHI'2004, pp 49–56
- Simmetrics Open Source Library <http://www.dcs.shef.ac.uk/%7Eesam/simmetrics.html>
- Thibault P, Bourgeois P, Hess U (2006) The effect of group-identification on emotion recognition: the case of cats and basketball players. *J Exp Soc Psychol* 42:676–683
- Traum D, Swartout W, Marsella S, Gratch J (2005) Fight, flight, or negotiate: believable strategies for conversing under crisis. In: Proceedings of the 5th international conference on intelligent virtual agents, Kos, Greece, pp 52–64