British Journal of Educational Technology *Vol 44 No 2 2013* 197–207
doi:10.1111/j.1467-8535.2012.01286.x

Modelling human emotions for tactical decision-making games

Gillian C. Visschedijk, Ard W. Lazonder, Anja van der Hulst, Nathalie Vink and Henny Leemkuil

Gillian Visschedijk, Anja van der Hulst and Nathalie Vink work at the Departments "Training & Performance Innovations" and "Military Operations," TNO Defence, Security and Safety. They specialise in serious games for all kinds of military training. Ard Lazonder and Henny Leemkuil work at the Department of Instructional Technology, University of Twente. They specialise in technology-enhanced learning and have a particular interest in learning with games and simulations. Address for correspondence: Ms Gillian C. Visschedijk, TNO Defence, Security and Safety, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands. Email: gillian.visschedijk@tno.nl

Abstract
The training of tactical decision making increasingly occurs through serious computer games. A challenging aspect of designing such games is the modelling of human emotions. Two studies were performed to investigate the relation between fidelity and human emotion recognition in virtual human characters. Study 1 compared five versions of a virtual character that expressed emotions through different combinations of posture, facial expression, and tone of voice. Results showed that emotion recognition was best when all three behavioural cues were present; posture + face and posture + tone of voice were joint second best. In study 2, these three versions were supplemented with contextual information. Cross-variant comparisons yielded marginal differences in emotion recognition and no differences in tactical decision making. Together, these findings suggest that the combination of posture with either facial expression or tone of voice is sufficient to ensure recognition of human emotions in tactical decision-making games.

Introduction

Tactical decision making denotes the ability to choose which actions or solutions should best be taken to accomplish a goal or task. The decisions that emanate from this process can literally be of vital importance to professionals such as police officers, fire fighters, security guards and military commanders who operate under dangerous or threatening conditions. The development of tactical decision-making skills increasingly occurs through serious games that, due to advanced computer technology, enable commanders-in-training to make tactical decisions in situations that are impossible in the real world for reasons of safety, cost and time (Kiili, 2007; Knerr, 2006).

Tactical decisions are made by matching features of the current situation to previously acquired patterns, selecting the most appropriate course of action and mentally simulating how these actions would work out in the current situation (Klein, 2008). The first step in the decision-making process thus hinges on situational awareness, which often requires the ability to recognise human emotions. Imagine, for instance, a riot police officer controlling a right extremists' demonstration. His decision to either intervene or refrain from action depends almost exclusively on his assessment of the crowd's level of aggression (Moya, McKenzie & Nguyen, 2008). Gaining proficiency in human emotion recognition thus seems to require maximum training fidelity: the more realistically human emotions are modelled in a game, the better they are recognised. Paradoxically, however, high-fidelity games are particularly expensive to develop while they do

Practitioner Notes

What is already known about this topic

- Tactical decision making often involves the recognition of human emotions.
- Humans exhibit their emotions through facial expressions, body movement and posture and tone of voice.
- Some of these behavioural cues are more predominant than others.
- Environmental cues convey important additional information to help recognise or infer emotional states.

What this paper adds

- Not all three behavioural cues need to be present to recognise human emotions in a known context.
- The combination of posture with either facial expression or tone of voice is sufficient to recognise human emotions.
- Both cue combinations lead to equally high recognition rates and qualitatively comparable tactical decisions.

Implications for practice and/or policy

- Designers of tactical decision-making games can lower the quality of the game characters' facial expressions or omit their vocalisations.
- As high-end graphics involve high development costs, the former option seems the most cost-effective.

not necessarily enhance learning (Feinstein & Cannon, 2002; Mania, Wooldridge, Coxon & Robinson, 2006). The present research therefore aspired to establish whether and how the degree of realism (ie., fidelity) in representing human emotions influences emotion recognition.

The modelling and recognition of emotions is key to a wide range of applications of virtual human characters. Examples include avatars in multi-user virtual environments such as Second Life and animated pedagogical agents in e-learning environments. Particularly noteworthy is the work of Baylor, who examined the design and effects of pedagogical agents on motivational and learning-related outcomes (eg., Baylor, 2011; Baylor & Kim, 2009; Kim, Baylor & Shen, 2007). However, even though the modelling of emotions is pivotal to this research, their recognition is assumed rather than assessed. The present research is thus complementary to Baylor's in that it involved a direct assessment of emotion recognition while using a similar interpretation of fidelity and research methodology.

Human beings have a wide range of emotions, which they exhibit mainly through facial expressions, bodily movements (mostly posture) and tone of voice (Argyle, 1988; Baylor, 2011). Whether these behavioural cues are intentionally communicative or not, they often suggest considerable information about a person's emotional arousal (Gratch & Marsella, 2001). Attempts to efficiently model emotional states in virtual human characters are often based on the cue dominance approach. This approach rests on the notion that some cues are more relevant than others, even if they represent the same information. The cues that are most relevant are called dominant cues; weaker cues are neglected if they appear simultaneously with dominant cues representing the same information (Warren & Riccio, 1985).

Cue dominance can be established from research. Various affective computing specialists have pointed out that facial expressions of virtual human characters tend to be somewhat ambiguous

(eg., Donath, 2001). Argyle (1988) conjectured that this ambiguity is reduced when facial expressions are accompanied with congruent bodily cues. This was substantiated by Vinayagamoorthy, Brogni, Steed and Slater (2006), who showed that posture was a more important indicator of a virtual character's emotional state than facial expression but challenged by Clavel, Plessier, Martin, Ach and Morel (2009) who found the opposite to be true. Tone of voice, the third behavioural cue, is omnipresent in multi-user virtual environments and can facilitate emotion recognition from facial cues (Sebe, Cohen, Gevers & Huang, 2006). However, Bailenson, Yee, Merget and Schroeder (2006) found comparable recognition rates in their "voice" and "voice + face" condition.

These inconsistent findings have given insufficient grounds to establish a cue dominancy hierarchy. A possible explanation is that the cited studies did not take the impact of environmental cues into account. Carroll and Russell (1996) asserted that contextual information is crucial to identify or infer emotional states. To illustrate, a different emotion is attached to a crying avatar in Second Life when the accompanying chat message explains that the individual whom the avatar represents has either lost his job or got promoted (Noël, Dumoulin & Lindgaard, 2009). It thus seems plausible that emotion recognition is sensitive to the meaning of the scene in which a virtual human character appears. This, in turn, implies that contextual cues should be taken into consideration in determining which behavioural cues are needed to adequately train human emotion recognition in tactical decision-making games.

The present research investigated this issue in two studies. Based on the cue dominance approach, it was assumed that maximum fidelity of all three behavioural cues may not be needed for an observer to recognise the emotional state of a virtual human character. Study 1 therefore sought to establish a cue dominance hierarchy by assessing the relative effectiveness of different combinations of behavioural cues to represent the six emotional states relevant for tactical decision making, which were derived from a task analysis in the fields of infantry tactics and crowd and riot control tactics (Visschedijk, 2010). In keeping with Baylor and Kim (2009), fidelity was defined by the mere presence or absence of behavioural cues, which was deemed more feasible than trying to define and compare distinct fidelity levels for each individual behavioural cue. Study 2 investigated whether this hierarchy would hold if emotion recognition occurs in context.

For the sake of experimental rigour, both studies were performed outside the context of a tactical decision-making game. While this admittedly lowers external validity, it increases internal validity in that research participants can be asked to recognise the exact same emotions in the exact same order. Such controlled conditions are very difficult to achieve in an actual game where the emotions shown by a game character depend on the player's actions. The implications of this choice of research setting are addressed in the general discussion.

Study 1

Method

Participants

Twenty-eight adult volunteers participated in this study. There were 16 males and 12 females with a mean age of 32 years. Participants were neither trained in tactical decision making nor in human emotion recognition.

Materials

The study used 30 computer animations of a virtual human character against a light grey background that were designed with Moviestorm (2005). The character could show an emotion through facial expression, posture or tone of voice. As the character's gender might influence how emotions are expressed or classified by the observer, the study used a single male character;

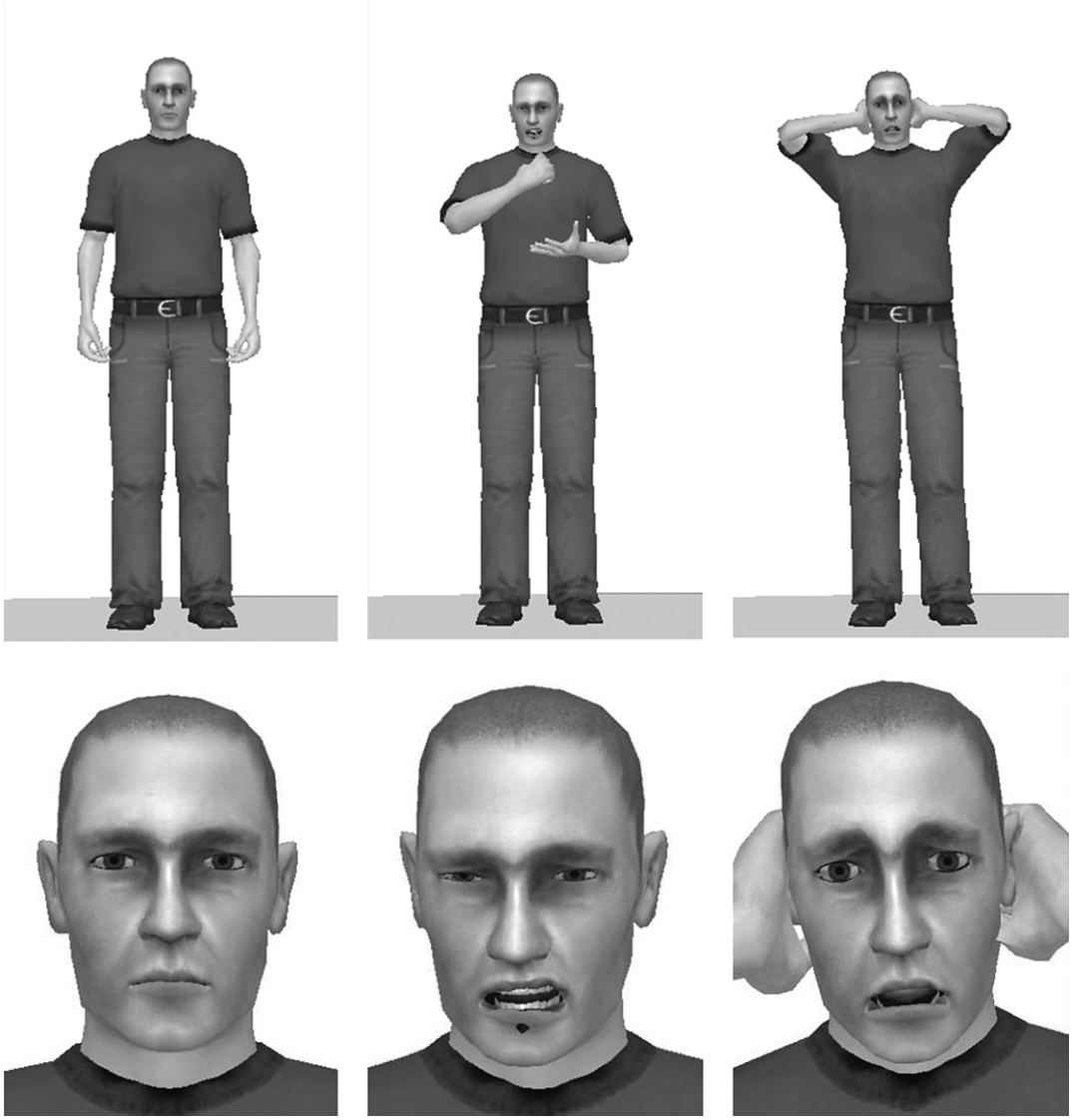

Figure 1: Still images of the virtual human character in a neutral (left), aggression (middle), and panic (right) emotional state. These stills are taken from the 5-second animations that were displayed on a 17-inch monitor (Study1) or via a beamer (study 2)

the choice of this gender was arbitrary. Six emotional states relevant to tactical decision making were modelled: neutral, anger, aggression, fear, panic and elation (see Figure 1 for examples). Facial and postural expressions of each emotion were taken from Moviestorm's library and checked for accuracy against existing standards (Coulson, 2004; Fabri, Moore & Hobbs, 2002; Gunes & Piccardi, 2007; Kleinsmith, De Silva & Bianchi-Berthouze, 2006). Tone of voice was added from Internet-retrieved crowd sound samples. Crowd sounds were preferred because they are prevalent in tactical decision-making situations and do not reveal the information an individual seeks to convey (which would facilitate the recognition of his or her emotional state). Draft versions of the 30 animations were subjected to a pilot test with four individuals who did not participate in the study. Final minor improvements were made based on their comments.

Five variants were created of each emotion that differed only with regard to cue presence. Variants were: posture (P), posture and facial expression (P + F), posture, facial expression, and tone of voice (P + F + V), posture and tone of voice (P + V) and facial expression and tone of voice (F + V). Variants without tone of voice had no sound at all, whereas omitted cues in variants without posture or facial expression were set to "neutral." This decision was made on practical grounds: in tactical decision-making games, it would be quite odd to leave out the rest of the body or remove the face.

Two possible variants, facial expression only and tone of voice only, were not included. The former was left out because facial expression alone was deemed insufficient to recognise emotions from a distance (Gunes, Piccardi & Pantic, 2008)—which is prevalent in tactical decision making; the latter because sounds such as laughs, cries or sighs are not uniquely linked to one single emotional state (Russel, Bachorowski & Fernández-Dols, 2003).

Procedure

Participants took part in the experiment one at a time. They were seated behind a laptop with a 17-inch monitor attached. Following a brief introduction, the participants were shown a still image of the human character with a neutral expression. It was explained that this appearance signalled the beginning and end of every animation and that the character *could* show an emotion in between for 5 seconds. The participants were left ignorant of which emotions they might encounter and were instructed to verbally describe the emotional state they recognised; the experimenter would write down their answers. The participants could repeat an animation as often as they liked before moving on to the next one. After these instructions, the participants watched and judged all 30 animations (the same six emotions modelled in five variants). The order of variants and emotions within variants was counterbalanced and participants were randomly allocated to one of four versions.

After data collection, the participants' responses were coded as true or false using a rubric that was developed from the reactions obtained in the pilot test and extended with synonyms from a dictionary. All responses were coded by two raters (Cohen's $\kappa = 0.72$); different codings were resolved through discussion.

Results

Table 1 presents the mean proportion of correct recognition by variant and type of emotion. Two-way repeated measures analysis of variance (ANOVA) showed that recognition of emotions

Table 1: Mean proportion of correct recognition by variant and emotion

	*Variant**				
	P	*P + F*	*P + F + V*	*P + V*	*F + V*
Study 1	0.51 (0.17)	0.69 (0.14)	0.78 (0.14)	0.64 (0.17)	0.58 (0.15)
Study 2	–	0.89 (0.15)	0.90 (0.13)	0.81 (0.11)	–

	Emotion					
	Neutral	*Anger*	*Aggression*	*Fear*	*Panic*	*Elation*
Study 1	0.96 (0.08)	0.76 (0.25)	0.64 (0.26)	0.30 (0.32)	0.46 (0.24)	0.71 (0.21)
Study 2	0.89 (0.23)	0.99 (0.07)	1.00 (0.00)	0.57 (0.33)	0.79 (0.32)	0.96 (0.11)

Note: standard deviations in parentheses.
*P = posture, P + F = posture + face, P + F + V = posture + face + tone of voice, P + V = posture + tone of voice, F + V = face + tone of voice.

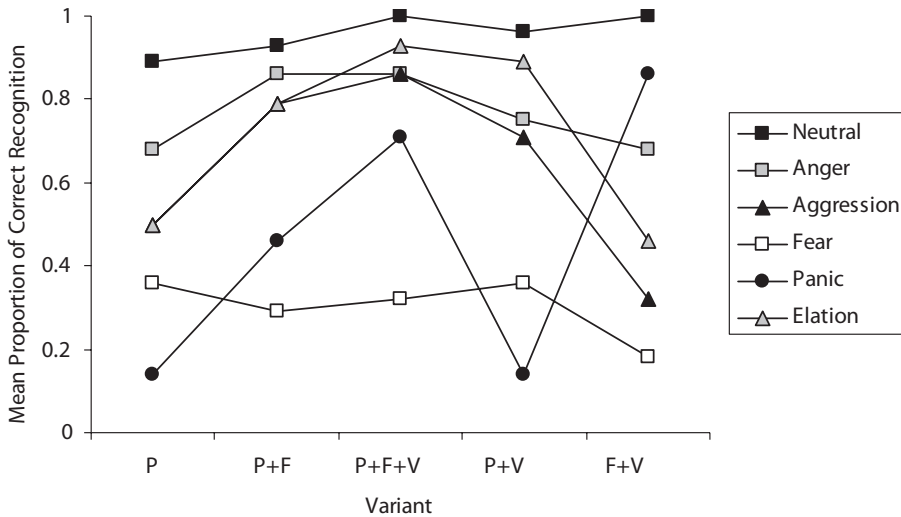

Figure 2: Mean proportion of correctly recognised emotions by variant (study 1)

overall differed among variants [$F(4,108) = 14.56$, $p < .01$] and that some emotions were more easily recognised than others [$F(3.67,99.17) = 25.73$, $p < .01$][1]. A significant variant × emotion interaction [$F(11.10,299.94) = 6.50$, $p < .01$] further indicated that these differences were not the same in each variant. Figure 2 visualises how recognition of all six emotions differed within each of the five variants. Post hoc comparisons, using paired-sample *t*-tests with Bonferroni correction ($\alpha = 0.005$), revealed that three emotions (neutral, anger and fear) were equally well recognised in all five variants [$t(27) < 1.80$, $p > .08$]. Data for the remaining three emotions paint a mixed picture. Recognition of "aggression" and "elation" depended on variant mainly because both emotions were relatively well recognised in the P + F + V variant [$t(27) > 3.38$, $p < .005$] and relatively poorly in the F + V variant [$t(27) = 3.10$, $p < .005$]. Results for "panic" were even more divergent and difficult to interpret. It nevertheless appears that recognition of this emotion depended heavily on facial expression and tone of voice.

Conclusion

This study sought to rank various combinations of behavioural cues according to their effectiveness to represent emotional states in a virtual human character. Recognition in general was most successful in the P + F + V variant where nearly eight out of 10 emotional states were correctly identified. The P variant had the lowest overall score with approximately one out of two emotional states recognised. However, a definitive ranking can not be made due to the significant interaction effect, and this complicates the choice of variants for study 2. The goal of study 2 was to assess whether contextual cues would influence the ranking. As not all variants could be included in study 2 for practical reasons, a more liberal interpretation of the results might help to decide which variants should be further investigated.

In doing so, the interaction effect was taken as a sign to reconsider the representation of emotional states that were disproportionally recognised across variants (ie., aggression, panic and elation). Although these emotional states might inherently be more difficult to recognise (Clavel *et al*, 2009; Crane, 2009; Vinayagamoorthy *et al*, 2006), their representation might be improved so as to decrease cross-variant differences in recognition. With "aggression" and "elation" in particular, these improvements could increase recognition in the P and F + V variant (see Figure 2), but it seems unlikely that the gap between these variants and the other three can be

entirely bridged. Assuming that this postulation holds true, it can be tentatively concluded from Figure 2 that most emotional states are best recognised in the P + F + V variant and that the P + F and P + V variants are joint second best.

This provisional conclusion is substantiated by planned contrasts among the variants (this datum was not reported in the results section because of the significant interaction effect but deemed informative for this specific purpose). It was found that the proportion of correct recognition in the P + F + V variant was significantly higher compared with all other variants. The difference between the P + F and P + V variant was not significant, while both variants had significantly higher recognition rates than the P variant. The F + V variant was less effective than the P + F variant (but not the P + V variant) and as effective as the P variant.

Based on these interpretative analyses, it was decided to include the P + F + V, P + F and P + V variant in Study 2. The first goal of this second study was to assess whether contextual cues could reduce cross-variant differences in recognition. A second goal was to find out whether contextual cues would cause all emotional states to be comparably recognisable in all variants.

Study 2

Method

Participants

Twenty-four students from the Dutch military police academy participated in this study. The sample was composed of all males with a mean age of 28 years. These students had some experience in tactical decision making and human emotion recognition and were typical of the target audience for tactical decision-making games.

Materials

Stimulus materials were the 18 animations from the P + F + V, P + F and P + V variants used in study 1. Variants were left intact except that every animation was supplemented with an introductory context description of the setting (the type of people involved, the area where they have gathered and the reason for their gathering), the assignment (the commander's order, such as "control the crowd and prevent escalation") and the penultimate event (eg., "some crowd members have just been arrested") (cf. Klein, 2008). The former two aspects are part of a standard briefing; the latter was added because emotions are often triggered by some preceding event that, unlike in tactical decision-making games, would otherwise remain covert. Six contextual descriptions were used in total, one for every emotion. The context descriptions were checked by four subject matter experts; minor adjustments were made based on their comments.

An answer form containing two open-ended questions was designed to gauge participants' interpretation of each animation. The first question asked the participants to describe which emotion the animation represented, the second question inquired after their commanding decisions in case the whole crowd would be in this emotional state.

Procedure

The study was conducted in one session that took place in a regular classroom. Instructions were similar to those of study 1 except that the participants were directed to fill out the answer form after each animation. The two questions were clarified and the participants were instructed to answer them as if they were a platoon commander. After the instruction, the experimenter read aloud the first context description and showed the associated animation twice via a beamer. The participants then had 2 minutes maximum to complete the answer form. This time limit was imposed for practical reasons (the experiment had to be performed within one lesson); experiences gained in study 1 proved that 2 minutes would be more than enough time to answer each question. The remaining 17 animations were administered similarly. Variants appeared in ascending order based on the recognition scores from study 1.

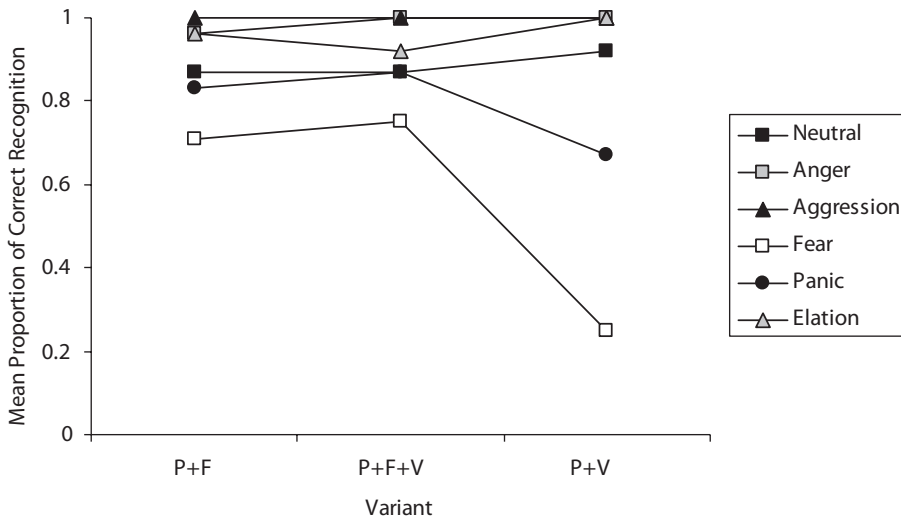

Figure 3: Mean proportion of correctly recognised emotions by variant (study 2)

The rubric from study 1 was used to score the participants' answers to the first question on the answer form. All responses were coded by two raters (Cohen's $\kappa = 0.80$); different codings were resolved through discussion. The participants' answers to the second question were judged against those of a tactical decision-making expert who had watched all animations while being knowledgeable of the virtual human's emotional state (Cohen's $\kappa = 0.85$). As some situations involved multiple decisions, the maximum score per variant was 13.

Results

The participants' emotion recognition rates (see Table 1) were analysed to reveal whether the presence of contextual cues could make up for the absence of behavioural cues. Two-way repeated measures ANOVA produced a main effect of variant [$F(2,46) = 5.53$, $p < .01$], a main effect of emotion [$F(2.84,65.31) = 13.91$, $p < .01$] and a significant interaction [$F(10\ 230) = 6.44$, $p < .01$]. As shown in Figure 3, recognition of four emotions (neutral, anger, aggression and elation) was comparable across variants; this was substantiated by paired-sample t-tests with Bonferroni correction [$t(23) < 1.46$, $p > .017$]. "Fear" and "panic" were less often recognised in the P + V variant compared with the other two variants. In case of "fear," these differences reached statistical significance [$t(23) > 3.82$, $p < .017$], in case of "panic," they did not [$t(23) < 2.46$, $p > .017$].

The participants' commanding decisions were analysed by one-way repeated measures ANOVA with variant as within-subject factor and the number of incorrectly recognised emotions as covariate. Results showed that commanding decisions in the P + F variant (mean [M] = 3.38, standard deviation [SD] = 1.84) and P + V variant ($M = 3.21$, $SD = 1.91$) were as good as those in the P + F + V variant ($M = 3.17$, $SD = 1.61$) [$F(2,44) = 0.09$, $p = .91$]. The covariate had no effect on these scores [$F(1,22) = 0.09$, $p = .77$], indicating that the quality of the participants' decisions was generally independent of their ability to recognise and describe emotional states. The non-significant variant × incorrect recognitions interaction [$F(2,44) = 0.08$, $p = .92$] further proved that this was the case in all three variants.

General discussion

Tactical decisions often rest on the recognition of human emotions. People express their emotions through multiple modalities such as their speech, their face and their body. This research supports

the prediction that a subset of these behavioural cues is sufficient to properly recognise an emotion in a known context. This conclusion has direct implications for the design of tactical decision-making games, and provides concrete directions for future research.

Results from study 1 confirm the notion that emotion recognition tends to improve with multiple behavioural cues (cf. Argyle, 1988; Clavel *et al*, 2009; Crane, 2009; Vinayagamoorthy *et al*, 2006). Recognition rates were highest in the P + F + V variant and lowest for the P variant; the differences between the three bimodal variants was less pronounced and depended on the type of emotion. Study 2 yielded superior recognition compared with study 1, but this result can not be attributed solely to the contextual cues because the military police academy students were more experienced in emotion recognition than the adult volunteers in study 1. This indistinctness is of minor importance here because the present research focused on tactical decision-making games that, by definition, contain contextual cues and are intended for trainees with some prior knowledge. It would nevertheless be interesting for future research to investigate the effects of contextual cues and subject knowledge in isolation. Study 2 further showed comparable recognition rates in the P + F + V and P + F variants. The P + V variant was as effective as the P + F variant but less effective than the P + F + V variant. These findings suggest that postural and facial expressions are dominant cues for recognising human emotions in known circumstances, whereas tone of voice is a weak cue with little added value.

However, suboptimal recognition in the P + V variant was mainly due to the emotion "fear"; the other emotions were as well recognised as in the P + F + V variant. This result implies that facial cues are dominant in the recognition of "fear" only. A possible explanation is that "fear," which is difficult to recognise anyway (Argyle, 1988; see also study 1), is mainly expressed through the face and associated with rather subtle body movements (slight crouching) and ambiguous sounds (soft, shivery breathing). The other emotions, by contrast, rely less heavily on facial expressions and have more explicit bodily movements and sounds that facilitate recognition in absence of facial cues.

More importantly, the quality of the participants' tactical decisions was comparable among variants. Even though the overall scores were somewhat low (which is quite understandable given that participants were tactical decision-making trainees), the least well-recognised emotions "fear" and "panic" showed substantial internal consistency. This result suggests that the relatively poor recognition in the P + V variant did not impinge on the eventual tactical decision. And because making accurate decisions is the ultimate goal of tactical decision-making training, it seems fair to consider the P + V variant as viable low-fidelity alternative to represent most human emotions.

Taken together, this research provides evidence that posture with either facial expression or tone of voice are dominant cue combinations that suffice to inform tactical decision making if trainees are briefed on the context and background of the target situation. Future research should reveal whether facial expression and tone of voice qualify as dominant combination as well because the F + V variant was not included in Study 2. This conclusion is generally consistent with the cue dominance approach (Warren & Riccio, 1985) and demonstrates that the results found in fundamental emotion recognition experiments extend to more realistic settings. The latter claim is supported by the fact that the present research addressed all emotions relevant to tactical decision making and assessed their recognition through a free response format (which is more valid than selecting emotions from a predefined list). Additionally, as this research went beyond the mere recognition of emotions by considering participants' tactical decisions, its results generalise to situations in which emotion recognition is a means rather than an end.

However, generalisability might be challenged by the use of a male character to express the emotions. Would the same results be obtained if a female character had been used? It seems

plausible that men and women express the same emotions differently, or with different intensity, and this might influence recognition. Likewise, similar expressions may be variously classified depending on the gender of the virtual character. A screaming woman, for instance, is presumably more often associated with the emotion "panic" than a screaming man. A related issue concerns the virtual character's clothing. While uniformly dressed in the present research, different outfits and headwear could either facilitate or complicate emotion recognition. In addition, demographic and cultural characteristics such as race, nationality and foreign accents might influence emotion recognition as well. Future research should therefore establish how a virtual character's appearance influences emotion recognition.

Future studies could also strengthen the validity of the present findings. One suggestion would be to compare the results from study 2 to emotion recognition and commanding decisions in an actual tactical decision-making game. The design of this research, although seemingly straightforward, will need some careful consideration because developing three versions of the same game is a rather expensive endeavour. Another interesting avenue for further research would be to replicate study 2 with a sample of decision-making experts. Their decisions would probably be more appropriate than those from the trainees in study 2. Whether their decisions will also be more consistent across variants remains to be shown.

Practical implications pertain to the design of digital tactical decision-making games. Due to their specific contents, these games are usually custom-made and not freely available. Their development costs are not made public either, but estimates are that their commercial equivalents (war games, strategy games) require over 20 million euros to develop (Prensky, 2008; Takatsuki, 2007). Serious game designers looking for ways to cut development costs while maintaining training effectiveness could either lower the quality of the game characters' facial expressions or omit their vocalisations. As high-end graphics typically involve high development costs, the former option seems the most cost-effective. Both recommendations might generalise to other serious game genres that hinge on human emotion recognition (eg., role playing games, medical simulation games) and the design of virtual coaches and pedagogical agents.

Note

1. As the sphericity assumption was violated, Greenhouse–Geisser corrected degrees of freedom are reported.

References

Argyle, M. (1988). *Bodily communication*. New York: Methuen.

Bailenson, J. N., Yee, N., Merget, D. & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence*, *15*, 359–372.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, *59*, 291–300.

Baylor, A. L. & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: when less is more. *Computers in Human Behavior*, *25*, 450–457.

Carroll, J. M. & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*, 205–218.

Clavel, C., Plessier, J., Martin, J. C., Ach, L. & Morel, B. (2009). Combining facial and postural expressions of emotions in a virtual character. In Z. Ruttkay, M. Kipp, A. Nijholt & H. H. Vilhjálmsson (Eds), *Intelligent virtual agents* (pp. 287–300). Berlin: Springer.

Coulson, M. (2004). Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*, 117–139.

Crane, E. A. (2009). Measures of emotion: how feelings are expressed in the body and face during walking. (Unpublished doctoral dissertation, The University of Michigan).

Donath, J. (2001). Mediated faces. In M. Beynon, C. L. Nehaniv & K. Dautenhahn (Eds), *Cognitive technology: instruments of mind* (pp. 373–404). Berlin: Springer.

Fabri, M., Moore, D. & Hobbs, D. (2002). Expressive agents: non-verbal communication in collaborative virtual environments. In A. Marriott, C. Pelachaud, T. Rist, Z. Rutthay & H. Vilhjalmsson (Eds.), *Proceedings of the first international joint conference on autonomous agents and multi-agent systems* (pp. 402–409). New York: ACM.

Feinstein, A. H. & Cannon, H. M. (2002). Constructs of simulation evaluation. *Simulation and Gaming*, *33*, 425–440.

Gratch, J. & Marsella, S. (2001). *Tears and fears: modeling emotions and emotional behaviors in synthetic agents.* Paper presented at the Fifth International Conference on Autonomous Agents, Montreal, Canada.

Gunes, H. & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, *30*, 1334–1345.

Gunes, H., Piccardi, M. & Pantic, M. (2008). From the lab to the real world: affect recognition using multiple cues and modalities. In J. Or (Ed.), *Affective computing: focus on emotion expression, synthesis, and recognition* (pp. 185–218). Vienna: Tech Education and Publishing.

Kiili, K. (2007). Foundations for problem-based gaming. *British Journal of Educational Technology*, *38*, 394–404.

Kim, Y., Baylor, A. L. & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, *23*, 220–234.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*, 456–460.

Kleinsmith, A., De Silva, R. & Bianchi-Berthouze, N. (2006). Cultural differences in recognizing affect from body posture. *Interacting with Computers*, *18*, 1371–1389.

Knerr, B. W. (2006). *Current issues in the use of virtual simulations for dismounted soldier training.* Paper presented at the NATO Human Factors & Medicine Panel Workshop on Virtual Media for Military Applications, West Point, NY.

Mania, K., Wooldridge, D., Coxon, M. & Robinson, A. (2006). The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *12*, 396–404.

Moviestorm (2005). *Moviestorm (Version 1.2.0.1) [computer software]*. Cambridge: Moviestorm Ltd. Retrieved November 11, 2009, from http://www.moviestorm.co.uk.

Moya, L. J., McKenzie, F. D. & Nguyen, Q. H. (2008). Visualization and rule validation in human-behavior representation. *Simulation and Gaming*, *39*, 101–117.

Noël, S., Dumoulin, S. & Lindgaard, G. (2009). Interpreting human and avatar facial expressions. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. Oliveira Prates *et al* (Eds.), *Human–computer interaction—INTERACT 2009*, Vol. *5726* (pp. 98–110). Berlin: Springer.

Prensky, M. (2008). Students as designers and creators of educational computer games: who else? *British Journal of Educational Technology*, *39*, 1004–1019.

Russel, J. A., Bachorowski, J. & Fernández-Dols, J. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349.

Sebe, N., Cohen, I., Gevers, T. & Huang, T. S. (2006). Emotion recognition based on joint visual and audio cues. In Y. Y. Tang, S. P. Wang, G. Lorette, D. S. Yeung & H. Yan (Eds), *Proceedings of the 18th international conference on pattern recognition* (pp. 1136–1139). Los Alamitos, CA: IEEE Computer Society.

Takatsuki, Y. (2007). *Cost headache for game developers*. Retrieved June 29, 2010, from http://news.bbc.co.uk/2/hi/business/7151961.stm.

Vinayagamoorthy, V., Brogni, A., Steed, A. & Slater, M. (2006). The role of posture in the communication of affect in an immersive virtual environment. In H. Sun (Ed.), *Proceedings VRCIA '06: ACM International Conference on Virtual Reality Continuum and Its Applications* (pp. 229–236). New York: ACM.

Visschedijk, G. C. (2010). The issue of fidelity: what is needed in 3D military serious games? (Unpublished master's thesis, University of Twente, Enschede, The Netherlands).

Warren, R. & Riccio, G. E. (1985). Visual cue dominance hierarchies: implications for simulator design. *Transactions of the Society of Automotive Engineering*, *6*, 937–951.