Adapting Psychologically Grounded Facial Emotional Expressions to Different Anthropomorphic Embodiment Platforms

Marco Paleari, Amandine Grizard and Christine Lisetti
Affective Social Computing Group
Eurecom Institute
Sophia Antipolis, France
{paleari,grizard,lisetti}@eurecom.fr

Abstract

Starting from the assumptions that human-ambient intelligence interaction will be improved by having more human-human like communications and that facial expressions are fundamental in human-human communications, we describe how we developed facial expressions for artificial agents based on a psychological theory developed by Scherer.
In this current article we describe briefly the psychological theory that we have chosen as well as some of the reason for adopting it. We then describe the two different platforms we used, the Cherry avatar and the iCat Robot with a particular focus on the robot. Finally we explore the steps needed to adapt the psychological theory to the two different platforms and we present some conclusions and future development.

Introduction

"In future Ambient Intelligence (AmI) environments we assume intelligence embedded in the environment, its objects (furniture, mobile robots) and in its virtual, sometimes visualized agents (virtual humans)".

In this kind of scenario rendering the communications and the transfers of information between the humans and the AmI devices as much effective and efficient as possible is a central issue. Looking at human-human interactions as objective not only guarantees to approach this goal of efficiency but also helps to improve the quality of the interaction in term of pleasantness and naturalness (Gratch & Marsella 2005; Picard 1997).

It is commonly argued that an important part of the communication when talking directly with others is through paralanguage (e.g. voice tone and volume, body language). Actually natural human-human interactions, as generally all social interactions, also involve communication of internal states such as the affective processes (Besson *et al.* 2004; Merola & Poggi 2004; Mehrabian 1971; 1972; Picard 1997).

Emotions are multimodal (Lisetti & Gmytrasiewicz 2002), they involve facial expressions, body gestures and postures, voice tone, prosody and volume as well as heart rate, skin conductivity and a huge number of other signals. In human-human communications facial expressions and vocal information are supposed to carry the maximum of information (Besson *et al.* 2004; Mehrabian 1971).

In this work we focus on affective facial expressions displayed by social agents. We selected a psychological model of the emotion appraisal proposed by Scherer (Scherer 1982; 1985; 1987; Leventhal & Scherer 1987; Scherer 2001) and we created new facial expressions following the guidelines given by his *Component Process Theory*.

Facial expressions have been developed on two different platforms: Cherry, an avatar developed by the team through Haptek technologies, that has previously been used on a mobile robot (Lisetti, Brown, & Marpaung 2002) and iCat, a robot-cat developed by Philips Research for human-robot interactions (van Breemen 2005).

Similar works have been previously published on this topic (Kaiser & Wehrle 2001; Paleari & Lisetti 2006a; Grizard & Lisetti 2006). This current work differentiates from those previous ones as it is done on modern interfaces and focuses more on the difference in the approaches taken by adapting the theory and on the different issues involved. Furthermore more details about the process taken to develop the expressions are given.

This paper is developed as follows: first, we will explain what Scherer's theory is and how it is possible to create dynamic facial emotional expression from it; second we will present the two platforms on which we have displayed facial expressions. We will finally present user studies we have conducted to evaluate believability and recognition rates.

Scherer's theory for emotional facial predictions

Among the different theories representing emotions, the ways they arise and the way they can be represented, there is one that is more appropriated to the framework we are developing: this theory is the *Component Process Theory* (CPT) developed by Scherer (Scherer 1982; 1985; 1987; Leventhal & Scherer 1987; Scherer 2001).

We chose this theory not only to generate facial expressions but also to define our user model and to simulate agent emotion generation. There are three main reasons for choosing this particular theory:

1. it considers complex representations of the emotions and

of the way they arise[1].

2. it considers emotions with their complex three levels (sensory motor, schematic and conceptual) nature[2] (Leventhal & Scherer 1987).
3. it addresses emotive multimodal expressions[3] and gives guidelines for developing both emotive expression generation and recognition (Scherer 1987).

Scherer's emotional process

According to Scherer's theory emotions arise, in humans, from a process of appraisal of the surrounding events with respect to the well being of the organism.

This sequential process is described by some parameters called sequential evaluation checks (SECs). Some of these checks are linked to changes in the facial expressions: these are *novelty, intrinsic pleasantness, conduciveness and coping potential* (Scherer 1987; 2001). This process is represented in figure 1. For each component an emotional response is expressed, that we will call in the following sub-expression; the sequence of all these emotional responses determines the individual's emotional facial expression. In this sense the emotional expression generation is dynamic.

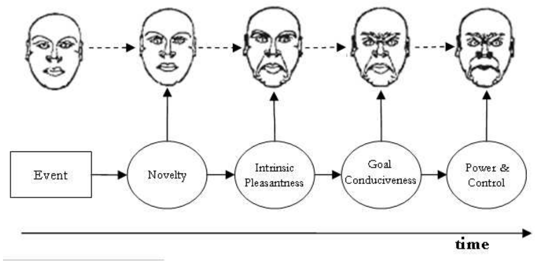


Figure 1: Computing representation of emotional facial process

The figure 1 has to be read as follows: when an event occurs, every person evaluates it in term of the novelty of the event; this evaluation is immediately linked to a reaction in terms of facial expression; the subject evaluates then the pleasantness (or un-pleasantness) of the event and the result of this evaluation is displayed with changes in the facial expression and so on and so forth.

Each sub-expression is coded with the Ekman's Facial Action Coding System (FACS) (Ekman, Friesen, & Hager 2002). This system codes facial expressions in term of minimal independent facial movements called Action Units (AUs) (Figure 2).

In the next subsection, we will explain how we used Scherer's theory to obtain emotional facial expressions.

[1]Emotions can be seen as modeled by a dimensional space with 16 different dimensions. The usual approaches are the 3-dimensional Pleasure, Arousal, Dominance space or the discrete emotions approach

[2]While other models usually considers only one level (e.g. the conceptual one)

[3]Facial expressions, voice emotive expressions and ANS emotive expressions.

Figure 2: Action Units Example (AU10, AU15, AU17, combination of the three)

	HAPPINESS	DISGUST
Novelty	low	open
Intrinsic Pleasantness	high	very low
Conduciveness	high	open
Coping potential	open	open
Normative Significance	open	open

Table 1: Predicted appraisal patterns for some emotions

Emotional Facial Expressions Generation from Scherer's theory

As mentioned in the former section, CPT gives informations to generate facial expressions associated to an occurred event.

While CPT links changes in facial expressions with the phases of the process of emotive appraisal of an event former approaches, in particular Ekman's (Ekman 1971; Ekman, Friesen, & Hager 2002), describes facial expressions as directely linked to the emotions themselves.

In particular CPT describes facial expressions as a dynamic process while the former approaches describe them as static (see (Scherer & H. 2006) for details).

An interesting point regards the possibility to easily display emotion mixture: since each emotion is represented differently, an emotion as happily surprised will just have a different SEC representation than happy or surprise (i.e. with high values for *Novelty* and *Pleasantness*) leading to a slightly different facial expression.

In table 1 we show some possible conversion between discrete emotional labels and their representations in terms of the SECs involved in facial expressions. Furthermore in table 2 we list the complete predictions as extrapolated from (Scherer 2001).

SEC	AUs
Novelty	1, 2 & 5
	4, 7, 26 & 38
Pleasantness	5, 26 & 38
	12 & 25
Unpleasantness	4, 7, 9, 10, 15, 17, 24 & 39
	16, 19, 25 & 26
Goal-Need Conduciveness (discrepant)	4, 7, 17 & 23
No Control	15, 25, 26, 41 & 43
	4 & 5
Control & High Power	7, 23 & 25
	23, 24 & 38
Control & Low Power	1, 2, 5, 20, 26 & 38

Table 2: SEC to AUs predictions (Scherer 2001)

To explain the process of generation of the facial expressions let us give the example of the discrete emotion “happiness”. For this emotion we have (tab. 1) *low* novelty which translates (tab. 2) in alternatively AUs 1, 2 & 5 or AUs 4, 7, 26; for the intensities all AUs will have the value ’b’ that on Ekman’s scale from ’a’ to ’e’ is a *low* value. Intrinsic pleasantness has high intensity: that translates into AUs 5, 26 & 38 or AUs 12, 25; the intensity is high and therefore ’d’. We replicate the same process for all the different components and one of the possible sequences of AUs (and intensities) we obtain is the following for happiness example:

- Novelty: AU1b, AU2b, AU5b
- Intrinsic Pleasantness: AU12d, AU25d
- Conduciveness: AU12c, AU25c

Two different platforms: a robot and an avatar

We are currently working on the creation of believable facial expressions based on the psychologically grounded approach described above for different embodiment platforms, and in particular on the Cleo robot and Cherry avatar which we now discuss.

Cleo robot

Cleo is the iCat robotic platform developed by Philips Research (van Breemen 2005) to study human-robot interaction. iCat has thirteen motors for facial expressions and body control (Figure **??**); four touch sensors and four multicolor LED are located in its feet and ears. Finally a webcam is hidden in its nose and speakers and microphones are integrated in its base. iCat has no legs and cannot move, therefore its main abilities are the social communicative ones.

iCat has been designed to be used as home companion for controlling an in-home network (ambient intelligence) or for gaming scenarios. Some simple application has been designed to test the platform: for example iCat has been an interface controlling a DVD recorder and a player in simple gaming scenarios (Saini *et al.* 2005; Bartneck, Reichenbach, & van Breemen 2004; van Breemen 2004a).

Philips research has created default facial expressions for iCat using principles of animations defined for Disney’s characters (van Breemen 2004b).

Our current research on Cleo aims at generating facial expressions based on a psychological theory as well as on cartoon animations, and comparing their believability to the expressions provided by Philips. Our goal is indeed to later link our psychologically-grounded facial expressions with the internal emotional states generated by using the same psychological theory as a model for designing computational and social intelligent architectures (Lisetti & Nasoz 2002).

We now describe the other embodiment platform that we used: Cherry, an anthropomorphic graphical avatar.

Cherry avatar

Cherry is an anthropomorphic avatar that can simply be inserted in applications and web pages (see Figure 4). We used tools developed by Haptek (hap 2006) which allow to design avatars belonging to both sexes from any human race (Lisetti, Brown, & Marpaung 2002). Simulation of different ages is possible with the use of morphs and skin textures.

Animation is based on a dedicated technology similar to MPEG-4 FAP (Facial Action Parameters). Different levels of control are available to move the avatars: from the control of global facial expressions, to morph and position of the avatar, to the control of basic facial movements.

Basic control of the avatar is possible by Haptek hypertext technology. Through hypertext one can, in fact, control text to speech, avatar position and launch Haptek switches which are collections of states representing still expressions of the avatar in term of combinations of some facial parameters. In other words through hypertext and switches one can control the evolution of the avatar expression over time as well as the softness of the transitions from one state to another.

We can imagine using this kind of avatar in home environments as interface to control the in-home network in the same ways we could do for the iCat robot or in simple e-tutoring (Paleari, Lisetti, & Lethonen 2005), e-marketing or gaming scenarios.

Adaptation of CPT to iCat

Scherer’s theory as been defined for humans. The use of this theory on a simple semi-anthropomorphic platform such Cleo need adaptation in term of semplifications .

Cleo has more limited expressive capabilities than humans; the first step is therefore to adapt the defined AUs to Cleo. In doing this, some AUs has to be ignored or extrapolated.

For example Cleo has one single servo to control the brows (AUs 1, 2 & 4); AUs 1 and 2 can be approximated with a turn of the brows outward (AU 1) or inward (AU 2) but AU 4 (brow lowerer) is not possible. We decided to express AU 4 in the same way of AU 2 (see table 3).

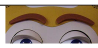
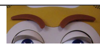
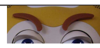
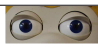
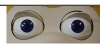
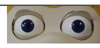
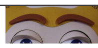
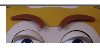
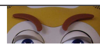
AUs	Neutral	Medium	Very High
SOME POSSIBLE ACTION UNITS			
AU2 (Outer Brow Raise)			
AU5 (Upper Lid Raiser)			
SOME EXTRAPOLATED ACTION UNITS			
AU4 (Brow Lowered)			
AU22 (Lip Funneler)			

Table 3: Some Possible and Extrapolated Action Units

With our created mapping between AUs and Cleo facial movements we designed emotional facial expressions based on Scherer’s theory.

With a previous user study, we acknowledged the need to improve the believability of these facial expressions (Grizard & Lisetti 2006). These results were explained by the limited degrees of freedom of Cleo’s face in terms of AUs compared to the richness of expressions of the human face.

We have adapted the animations by adding all Cleo possible bodily expression capacities: head, body, and eyes movements, as well as lights in paw and ears. In a second step we have exaggerated the intensity of iCat expressions using a similar approach to the one of Philips.

CPT describes head and eye movements giving guidelines (e.g. gaze: directed or aversion). We exploited these guidelines following, when needed, principles of animations (van Breemen 2004b). We, then, added the light (e.g. red color in the expression of anger) following CPT and simple social rules (red as caused by increased hearth rate and stroke or green for continue, ok, positive and red for stop, negative). We also exploited light pattern: for example we used a discontinuous and asymmetric light pattern for fear to address panic.

We would like to specify that the expressions we have displayed on Cleo follow the informations given in CPT. Each sequences of AUs are represented in a sequentially way, following and respecting SECs. However, in order to obtain believable and fluent facial expressions, some AUs can begin in SEC before or can finish in SECs after (by decreasing their intensities) the ones in which they are expressed.

Figure 3 shows the nine expressions we have designed using the CPT and the cartoon animation principles.

Figure 3: Designed Expressions for iCat.
Left to right and top to bottom: disgust, anger, fear, pride, contempt, happiness, sadness, indifference, shame

Adaptation to Cherry

Haptek avatars are very believable and the main issues arising in linking them to the CPT was not the lack of control over the facial expression (as with iCat) but rather the lack for some details.

In particular, we need information involving timings and intensities of the AUs concerned in the various component-related sub-expressions (see Figure 6).

Figure 4: One possible evolution of the expression for Fear according to Scherer's theory

Additionally some predictions seemed to advocate the activation of the same facial muscles [4] by two different components at the same time, leading to questionable believability of the expressions.

Finally we acknowledged that a wrong choice of the prediction (e.g. AUs 1, 2 & 5 instead of AUs 4, 7 & 26 for novelty) can lead to problem in terms of believability (see Figure 7).

To address some of these issues, we observed videos of the facial expressions performed by actors and performed several little adaptations:

- When possible we selected contiguous predictions that had the most AUs in common[5];
- We identified SEC that were shown by the actors longer and in a clear way and we augmented the duration and the intensities of the related AUs;
- We solved ambiguities and concurrences of activation of facial muscles by several AUs;
- We let the AUs fade out gently;
- We included head movements and basic avatar gaze (See (Paleari & Lisetti 2006a) for some other details about the process);

One example of the evolution of the resulting expression can be seen in Figure 4. All the designed expressions can be seen in Figure 5

Recognition and believability of our displayed expressions

We have conducted two user studies in order to evaluate believability and recognition rates of our new facial expressions. In the following section, we will present the protocols we have used and the results of these studies.

Recognition of Artificial Expressions by Humans

We have conducted a preliminary user study to first evaluate the recognition rate of the facial expressions we developed for the Cleo robot and Cherry avatar.

The experiment involved sixteen participants, four women and twelve men between twenty and thirty years old.

[4] For example AU 4 (brow lowerer) and AUs 1 and 2 (inner and outer brow raiser) are often activate sequentially by different components (e.g. novelty and intrinsic pleasantness)

[5] for example while selecting AUs 4, 7, 26 & 38 for *novelty* we selected, in case of un-pleasantness, the prediction identified by AUs 4, 7, 9 etc.

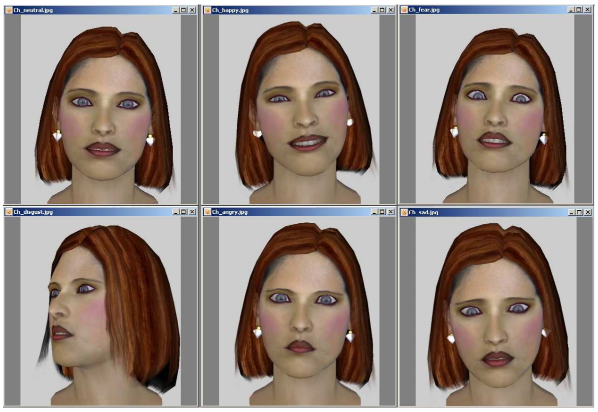


Figure 5: Designed Expressions for Cherry.
Left to right and top to bottom: neutral, happiness, fear, disgust, anger, sadness

They were asked to recognize the following expressions: happiness, disgust, contempt, sadness, pride, fear, anger, indifference and shame [6]. For this, each expression was shown twice and participants chose one expression among the nine (resp. five) possible expressions for Cleo (resp. Cherry).

We can conclude from the results, shown in (Grizard & Lisetti 2006; Paleari & Lisetti 2006a), that expressions were well recognized: 52% of accuracy for iCat and 93.8% for Cherry (but on the much simpler task of recognizing one expression out of five possible).

Believability Assessment of Artificial Expressions

In a second user study developed as the first one, the participants were asked to rate believability and exaggeration of the facial expressions. Ratings could be on the scale from 0 to 5. We displayed three different animations per emotion to the participants.

For iCat the three representations were the Philips implementation, the first version we implemented (direct porting of Scherer's theory without adaptation) and the second, adapted, version of our implementation.

For Cherry we presented animations of the avatar set with the default parameters from Haptek, the same avatar set with our, adapted, implementation and videos of an actor showing the facial expressions.

The results (Grizard & Lisetti 2006; Paleari & Lisetti 2006a) shown that in general the developed facial expressions demonstrate to be considered as believable. The result of our animations in term of believability are similar to the ones scored by the other representations (in some cases our animations are even considered as being more believable).

Exaggeration was rated in a similar way for all the different animations.

Open Research Questions

This work open some research questions and in particular:

- Are all the SEC related sub-expressions equally important? If they are not should some of them last longer or have more intensity?
- How should different sub-expressions fuse? (Figure 6)

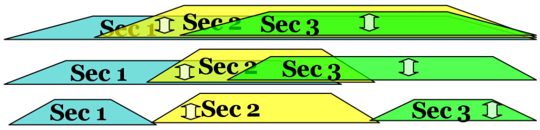


Figure 6: Timing and Intensities issues

- Are all AUs related to the same sub-expression equally important?
- How should we chose among the different possible predictions? (Figure 7)

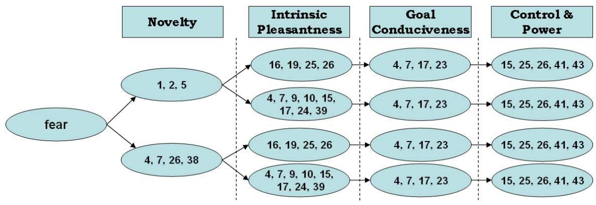


Figure 7: The multiple prediction issue for fear

- How can we consider other affective phenomena such mood?

Conclusion

The main objective of this current work is to design believable emotional facial expressions. For this we have chosen to use Scherer's psychological theory blended with cartoon animations for Cleo robot and completed by information from human videos for Cherry avatar.

We would like to precise that we had chosen Scherer's theory because it describes links between emotion, cognition and affective multimodal expressions. The main advantage of using such a theory for generating facial expressions is linked to the possibility of dynamically generating dynamic facial expressions for all the possible different emotions. With the CPT we do not need to design specific animations for specific emotions, the agent could directly convert the appraisal process into dynamic facial responses.

Another advantage of this theory resides into the possibility of using the same theory for both generation and recognition; in fact we can link AUs to SEC as well as we have linked SEC to AUs. Furthermore the same theory can be used to define the user model of emotions as well as it can be used for simulating the process of appraisal and therefore to simulate the generation of affective states belonging to the agent itself.

In (Paleari, Lisetti, & Lethonen 2005) we have shown a possible application of these facial expressions in a simple e-tutoring scenario. We are currently working at the developement on a simple interactive gaming scenario with Cleo.

[6]Only five expressions have been developed and tested on Cherry: happiness, disgust, fear, anger and sadness

Other current work includes the development of new facial expression for Cherry, the development of a Multimodal Affective User Interface (Lisetti & Nasoz 2002) including the multimodal fusion paradigm described in (Paleari & Lisetti 2006b) and the development of modules for automatic generation and recognitions of respectively AU sequences from SECs and SECs from recognized AUs.

References

Bartneck, C.; Reichenbach, J.; and van Breemen, A. 2004. In your face, robot! the influence of a character's embodiment on how users perceive its emotional expressions. In *Fourth International Conference on Design & Emotion*.

Besson, C.; Graf, D.; Hartung, I.; Kropfhusser, B.; and Voisard, S. 2004. The importance of non-verbal communication in professional interpretation.

Ekman, P.; Friesen, W. V.; and Hager, J. C. 2002. *Facial Action Coding System Investigator's Guide*. A Human Face.

Ekman, P. 1971. Universals and cultural differences in facial expressions of emotion. In Cole, J. K., ed., *Proceedings of Nebraska Symposium on Motivation*, volume 19, 207–283. Lincoln (NE): Lincoln: University of Nebraska Press.

Gratch, J., and Marsella, S. 2005. Lessons from Emotion Psychology for the Design of Lifelike Characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)* 19(3-4):215–233.

Grizard, A., and Lisetti, C. L. 2006. Generation of facial emotional expressions based on psychological theory. In *Working notes of 1rst Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence, June, 14-19, 2006, Bremen, Germany*.

2006. Haptek website (2006): www.haptek.com.

Kaiser, S., and Wehrle, T. 2001. *Appraisal processes in emotions: Theory, methods, research*. New York: Oxford University Press. chapter Facial expressions as indicators of appraisal processes, 285–300.

Leventhal, H., and Scherer, K. R. 1987. The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion* 1:3–28.

Lisetti, C., and Gmytrasiewicz, P. 2002. Can a Rational Agent Afford to be Affectless? A Formal Approach. volume 16, 577–609. Taylor & Francis.

Lisetti, C. L., and Nasoz, F. 2002. Maui: a multimodal affective user interface. In *Proceedings of the ACM Multimedia International Conference 2002*.

Lisetti, C. L.; Brown, S.; and Marpaung, A. 2002. Cherry, the little red robot with a mission and a personality. In *Working Notes of the AAAI Fall Symposium Series on Human-Robot Interaction*.

Mehrabian, A. 1971. *Silent Messages*. Belmont, CA: Wadsworth Publishing Company, Inc.

Mehrabian, A. 1972. *Nonverbal Communication*. Chicago: Aldine-Atherton.

Merola, G., and Poggi, I. 2004. Multimodality and gestures in the teachers communication. *Lecture Notes in Computer Science* 2915:101–111.

Paleari, M., and Lisetti, C. L. 2006a. Psychologically Grounded Avatar Expressions. In *Working notes of 1rst Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence, June, 14-19, 2006, Bremen, Germany*.

Paleari, M., and Lisetti, C. L. 2006b. Toward Multimodal Fusion of Affective Cues. In *Working notes of 1rst International Workshop on Human Centered Multimedia at ACM Multimedia 2006, October, 2006, S. Barbara, California*.

Paleari, M.; Lisetti, C. L.; and Lethonen, M. 2005. Valerie: a virtual agent for a learning environment, reacting and interacting emotionally. In *Working Notes of AIED 2005, 12th International Conference on Artificial Intelligence in Education*.

Picard, R. 1997. *Affective Computing*. Cambridge (MA): MIT Press.

Saini, P.; de Ruyter, B. E. R.; Markopoulos, P.; and van Breemen, A. J. N. 2005. Benefits of social intelligence in home dialogue systems. In Costabile, M. F., and Paternò, F., eds., *INTERACT*, volume 3585 of *Lecture Notes in Computer Science*, 510–521. Springer.

Scherer, K. R., and H., E. 2006. Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *(in press) Emotion*.

Scherer, K. R. 1982. Emotion as a process: Function, origin and regulation. *Social Science Information* 21:555–570.

Scherer, K. R. 1985. Emotions can be rational. *Social Science Information* 24(2):331–335.

Scherer, K. R. 1987. Toward a dynamic theory of emotion: The component process model of affective states. Geneva Studies in Emotion and Communication, 1(1), 1-98.

Scherer, K. R. 2001. *Appraisal processes in emotion: Theory, methods, research*. New York, NY, US: Oxford University Press. chapter Appraisal Considered as a Process of Multilevel Sequential Checking, 92–120.

van Breemen, A. 2004a. Animation engine for believable interactive user-interface robots. In *IROS 2004, 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

van Breemen, A. 2004b. Bringing robots to life: Applying principles of animation to robots. In *Proceedings of Shapping Human-Robot Interaction workshop held at CHI 2004*.

van Breemen, A. 2005. icat: Experimenting with animabotics. In *AISB*, 27–32.