Using Human Acting Skill to Measure Empathic Ability in Heterogeneous Characters

Bill Tomlinson
ACE (Arts Computation Engineering) program
University of California, Irvine
430A Computer Science Building
Irvine, CA, USA 92697-3425
Tel: +1-949-824-9333 / Fax: +1-949-824-4056
wmt@uci.edu

Abstract

This paper presents a method for evaluating the relative empathic ability of different characters. The method is based on the premise that human actors are able to give more compelling performances when acting opposite a more empathic acting partner. The method involves causing one or more human actors to engage in a short interaction with each character, videotaping these performances, showing the recordings to skilled acting instructors and asking these instructors to evaluate the quality of each performance. By controlling for the effects of order of presentation and other potentially confounding variables, we expect to be able to determine which of the characters are the most effective elicitors of high-quality performances from the human actors, and thereby to demonstrate the relative empathic ability of the various characters.

Keywords

Empathic agents, synthetic characters, acting, evaluation.

1. INTRODUCTION

Many different synthetic characters have been created by research and industry over the last several decades. Each of these characters has a different set of strengths, such as life-like skin, flowing hair, seamless movement, intelligent action-selection or emotional depth. However, there are few viable ways to compare these very different characters to each other. While it might be possible to compare some specific feature of several characters, it is much more difficult to assess the overall empathic ability of a given character. This paper proposes an experimental method for comparing the relative empathic ability of different characters.

The method presented here is based in a phenomenon found among dramatic actors, who find that they perform better when acting opposite other skilled actors. Stan Winston, head of Stan Winston Studios and maker of many of Hollywood's most famous animatronic characters (The Terminator, the dinosaurs from Jurassic Park and many others) summarized it as follows: "Any great actor will tell you that 50% of acting is reacting, and the greatest actors will also tell you that they give their finest performances when acting opposite another great actor. Imagine having to give your finest performance acting against nothing – a blue screen." [1] Blue or green screens are used when a computer animated character, synthetic background or other dramatic element is to be composited into a scene [2]. Animatronic characters, which actually appear on the film set, can be superior to computer-animated characters, which will be composited into the scene later, because the animatronics give a human actor something real to interact with while performing the scene. While skilled human actors are acknowledged to be the best acting partners, real-time characters such as animatronics are nevertheless considered superior to a blank screen when it comes time to perform a scene.

Stotland defines empathy as "an observer reacting emotionally because he perceives that another is experiencing or about to experience an emotion." [3] This definition resonates with the craft of acting, in which a performer reacts emotionally (or appears to react emotionally) [4] because he perceives that another actor is experiencing (or appearing to experience) an emotion. While there is certainly a difference between experiencing an emotion and appearing to experience an emotion, this difference does not undermine the ability of skilled actors to thrive when acting opposite skilled partners. Whether the actor's partner is actually feeling the emotion or simply reproducing the bodily signals of that emotion, the presence of appropriate emotional stimuli help the actor respond in kind. Actors therefore need to be (or at least appear to be) empathic (or perhaps empathogenic, that is, capable of eliciting empathy) while practicing their craft.

Given that not all "acting partners" are created equal, is it possible to measure which entities make the best acting partners? (The characters and other entities to be evaluated will be referred to as "acting partners" throughout this document.) This paper presents a method for comparing heterogeneous characters based on their empathic ability to elicit strong performances from human actors. The assumption on which the method is based is that actors will give better performances when acting opposite more empathic acting partners.

The method involves the following key steps:

- Videotape one or more human actors performing the same short dramatic scene opposite several different acting partners. Insure that the videotape

shows only the human actor, and does not reveal the acting partner.

- Show the videotaped performances to skilled acting instructors, and ask them to evaluate the quality of the performances.
- Analyze the results to determine which of the acting partners elicited the highest quality performances from the human actors.

Since the acting instructors will see only the human actors, this method makes it possible to compare characters as different as real humans, synthespians, autonomous robots, animatronic characters and videotaped actors using a consistent metric.

This paper describes a process by which heterogeneous entities may be evaluated for their ability to convey emotion and thereby elicit emotional responses (or the appearance of emotional responses) from human actors. The value of this work lies not just in making characters who are compelling acting partners in a few limited scenes. Rather, the key value lies in having a uniform way of measuring the empathic ability of a range of different characters; extrapolating from these experimental cases can help guide the production of empathically capable characters that may be of use across a much wider variety of domains.

2. RELATED WORK

There are several areas of previous work that relate closely to the design of this research methodology. The success of synthetic character research provides much of the inspiration and incentive for attempting to measure and compare vastly different characters. The area of human-computer relationships informs the work, since relationships, emotion and acting ability are all closely interlinked. Other researchers in numerous different fields have also shown that computers may have a positive effect on human performance. The broader area of dramatic acting provides much guidance in the craft of performing compelling scenes. Finally, work in film and special effects are relevant, since these fields have encountered (and tried to work around) some of the shortcomings of acting opposite a range of stimuli.

There is a long history of research focusing on the creation of synthetic characters. Most of this work has focused on interactive virtual humans [5]. Ken Perlin and his colleagues in the Media Research Lab at NYU have done pioneering work in creating synthetic actors (e.g., [6]). By working closely with the natural style of their characters' motion, they have created virtual characters who move and interact very naturally. Bruce Blumberg's Synthetic Characters Group at the MIT Media Lab has contributed in several areas including action-selection [7] and learning [8]. Justine Cassell's Gesture & Narrative Language research group at the MIT Media Lab builds virtual humanoids with the ability to express themselves like real people, in particular in the area of embodied conversational agents (e.g., [9, 10]). Through both the Oz Project at Carnegie Mellon University and the company Zoesis, Joseph Bates and his colleagues have created computational characters who appear lifelike and are able to interact with people in real time [11, 12]. More recently, Harald Schaub and his colleagues have presented research on empathic characters [13]. Helmut Prendinger and his colleagues offered an evaluation mechanism for empathic characters based on skin conductivity and other physiological signals [14]. Robots have also been endowed with characteristics that allow them to engage people in expressive ways [15, 16]. Many other researchers have also contributed to the creation of compelling synthetic characters, e.g., [17-24].

Various researchers have studied the relationships or potential relationships among people and computers. In their book "The Media Equation," Byron Reeves and Cliff Nass present an assortment of evidence that people respond to various forms of media in social and emotional ways [25]. Batya Friedman and her colleagues showed that people are very willing to engage in relationships with their AIBO robots [26]. In previous work, Bill Tomlinson has offered an emotion-based mechanism for computational entities to form social relationships with each other and with people [27].

Numerous researchers have shown that computer technologies may increase human performance. For example, a great portion of the CHI and CSCW conferences are devoted to the demonstration that human performance may be enhanced by various technological innovations. Sherry Turkle discussed how causing computational systems to function as companions, rather than tools, may enhance people's performance and quality of life [28]. Given this ongoing trend that points to the abundant ways in which computational systems may improve human performance, we believe that it is possible for well designed empathic characters to enhance human acting skill.

The study and practice of dramatic acting is central to the methodology proposed in this document. The connection beween acting and emotion has been discussed by acting theoreticians for centuries. In his essay "The Paradox of Acting", Denis Diderot offered that actors should not actually feel the emotion that they are playing, but should rather act "from thought," consciously imitating emotions rather than feeling them [29]. Method acting, on the other hand, proposed initially by Konstantin Stanislavky and popularized in America by Lee Strasberg, proposes that actors should strive to feel the emotions that they are portraying [4]. In his book "Masking Unmasked: Four Approaches to Basic Acting", Eli Simon recommends that, when acting with masks, actors "[a]llow emotions to wash through you; blocked feelings stymie honest reactions." (p. 33) [30] Whether the emotion of a scene comes from the actor actual feeling it or rather from a capable simulation of it, there is nevertheless an intimate tie between emotion and acting.

The art of acting for film requires a number of special techniques beyond those of stage acting [31], in particular for films relying heavily on special effects or fantastic characters. Film and video shoots frequently ask actors to perform in front of blue or green screens, into which their synthetic acting partners will later be composited [2]. This style of film making may cause difficulty for the actor, in

that he or she does not have a skilled actor to perform opposite, does not have a place to look, and does not necessarily even know what his or her acting partner looks like. Various ways to make the actor's job easier include having someone read the relevant lines of the missing actor and aiming a laser pointer at the approximate location where the actor should focus. This gives the actor the bare essentials to be able to perform the scene. However, the difficulty that actors have in performing opposite blanks screens points to the dynamic range of performances that these actors produce opposite various eliciting acting partners.

3. METHODOLOGY

The methodology proposed here involves using human actors and acting instructors to evaluate a wide range of different characters. By basing the evaluation mechanism on the action of an entity other than the character itself, certain elements that differ between characters, such as beauty or amount of motion, are largely factored out.

3.1 General Study

Prior to conducting the evaluation, the entities to be evaluated must be selected. These entities may take any form – robots, interactive characters, pre-recorded animated characters, human actors, videotaped actors, fixed images, a simple red dot at eye level, or a blank screen. The entities to be evaluated may also be accompanied by control-case entities, so that the results from this experiment might be compared to other experiments conducted previously or in the future.

The entities – both experimental and control – should then be standardized as much as possible. Size, audio volume, brightness and other factors should all be kept approximately constant. This standardizing will reduce the likelihood that one character will be more compelling due to factors that are not integral to the character itself.

The characters then need to be enabled to exhibit some form of common behavior. For a group of language-capable entities (for example, embodied conversational agents and human actors), this behavior might be to say a certain line of dialog from a play or movie. For non-language-based characters, the behavior might be to exhibit a certain emotion – for example, happiness, sadness, fear, disgust, anger or surprise [32]. For certain control conditions (e.g., a blank screen), it will of course not be possible for it to exhibit the behavior.

Once the acting partners have been prepared, human actors should be recruited to participate in the experiment. (We are planning tests to determine if the skill level of the human actor makes a difference to the results.) Each of these actors should be shown into the study room, and asked to stand at a certain location. He or she will then be presented with each of the acting partners in turn, and videotaped while he or she performs a short dramatic exchange with that partner. Each acting partner should be presented several times in a random order, so that the effect of the order of presentation is minimized.

When videotaping the actors' performances, it is important that no information be included that might reveal which character elicited the stimulus. If the evaluators are able to determine which stimuli elicited the performances, they could bias their evaluations in favor of certain acting partners rather than evaluating the quality of the performances themselves. Figure 1 shows the layout for the experiment.

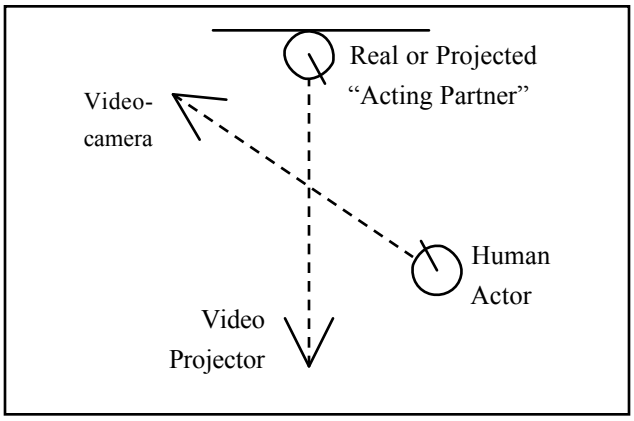

Figure 1: Experimental layout for recording human actor without revealing which "acting partner" elicited the performance.

Once the performances have been videotaped, they should each be edited so that they do not include the speech act of the acting partner, but rather only feature the response of the human actor.

These shots should then be randomly reordered, and presented to one or more skilled acting instructors. We choose to show the recordings to skilled acting instructors, rather than average people, because the skilled instructors have spent years discriminating among various closely related qualities of performance, helping their students hone their craft.

Each acting instructor should preview all of the performances, so that he or she has a sense for the overall quality range of the actor's performances. On a second viewing, the instructor should evaluate each performance with respect to a number of attributes, each on a six point Likert scale. Specifically, the instructor could be asked to evaluate the performances with regard to expressiveness, believability and technical proficiency. We chose an even number of Likert values so that the instructors do not have the option to give a performance a middle score, thereby forcing at least a binary choice.

Once the scores for all of the performances are recorded, they should be analyzed for several features. First, they should be analyzed to determine if there was a significant time effect (e.g., early performances were better than later performances or vice versa), either within the performances elicited by a single acting partner, or else across all acting partners. Second, they should be analyzed to determine which acting partners elicited the highest quality performances, controlling for any time effects. The corrected quality of performance (in particular regarding expressiveness and believability) should demonstrate any significant differences in empathic ability among the acting partners.

3.2 Specific Study

In the Social Code Group in the ACE program at UCI, we have a study currently under way using this methodology. In the study we are conducting, we are comparing the following acting partners:

- a human actor
- a videotape of the human actor
- an animated character with the voice of the human actor
- a still image of the animated character with the voice of the human actor
- a blank screen with the voice of the human actor.

The human actors will be drawn from a range of experience levels: first-year acting undergraduates, final-year acting undergraduates, first year MFA acting students and final year MFA actors. All students will be drawn from the UCI Drama Department.

The dialog we will be using is the following lines from Shakespeare's Macbeth (Act V, Scene V) [33]:

SEYTON: *The queen, my lord, is dead.*

MACBETH: *She should have died hereafter.*

The acting partner will play Seyton, while the human actor will play Macbeth. This short exchange was suggested by Robert Cohen, Claire Trevor Professor of Drama at UCI, as an effective segment through which to give actors a range of possible emotional responses. The fact that the lines are taken from a well-known play should make the study easier for the actors who will participate, as they will almost certainly know the back story of these lines.

While the simplicity of using just two lines of dialog may seem to simplify the problem too greatly, this very simplicity is one of the strengths of the approach. First, directors and acting coaches will sometimes spend a long time working on a very small chunk of dialog to encourage actors to focus their attentions. Actors should be capable of expressing their motivations and emotions in a very short period of time. Second, the brevity of the scene makes it more likely that various autonomous entities could be enabled to participate in future studies; building an entire suite of behaviors and dialog may be too great an effort for a production team to undertake for the sake of evaluation. Finally, the simplicity of the two lines allows this project to focus on the evaluation methodology itself, rather than spending a great deal of effort on building complex interactive characters at this point in the research. Once the methodology has been verified as potentially useful, the research team will build more complex characters that will serve as acting partners for future iterations of the study.

Each acting partner will be presented to the subject three times, for a total of fifteen performances. The video clips will then be shown to skilled acting instructors including Robert Cohen and Eli Simon, Associate Professor and Head of Acting at UCI, who will be asked to score them by quality of performance.

This experiment will explore three hypotheses:

- Actors will give the best performances opposite human actors, the worst opposite blank screens, and middle-range performances with video, animated characters and still images.
- More experienced actors will show a narrower range of performance quality with the varying acting partners and over repeated performances.
- Actors will give the best performances in the middle of the session, with the first performances being weaker because of a lack of practice, and the last performances being weaker because of excessive repetition.

Each of these three hypotheses pertains either directly or indirectly to the assessment of empathic ability in the entities serving as acting partners.

The first hypothesis is the core premise of the study – that different acting partners will elicit predictably different levels of performance. As mentioned earlier, the craft of acting is tied to the expression and elicitation of emotional responses. Stotland's definition of empathy involves two main parts – an entity *reacting* emotionally, and this reaction occurring because that entity *perceived* another's emotion. This study measures an entity's ability to elicit empathy from a human, which in turn represents the entity's ability to express emotion in a way that could be perceived by the actor.

If the second hypothesis holds true, then the study would suggest that less experienced actors would be more effective in evaluating the empathic ability of various acting partners, since their wider dynamic range would amplify the differences among the entities.

With regard to the third hypothesis, the possibility that actors may exhibit a changing level of performance over the course of the fifteen runs is indirectly relevant to the assessment of the empathic ability of the acting partners. This change could be a confounding effect that will need to be factored out before the effect of the differences caused by the acting partners themselves may be seen.

We have received approval from the Institutional Review Board (UCI IRB HS# 2004-3526) to conduct this study, and will be completing it over the next two months. The study will be completed prior to the AAMAS workshop, and I will be able to present further results from the study at that point.

4. DISCUSSION

There are numerous elements that helps a character to be effective as an empathic acting partner. The ability to express emotions, to make eye contact and to respond rapidly to an acting partner's behavior all help a character to elicit strong performances from a partner. All of these contribute to the overall empathic ability of a character.

One of the challenges that this experimental method makes clear is the value of interactivity in empathy. While a static character may provoke a short-term emotional response, an interactive component to the character is neces-

sary in order to sustain longer-term empathic relations. In longer scenes, interactive characters would have a clear advantage over pre-recorded characters.

This method is not without its drawbacks. If a language-based performance is used, it requires each of the characters to speak the appropriate lines of dialog. It requires all other elements of the scene to be held as constant as possible – cinematography, background sound, etc. Finally, certain characters may be more adapted to one style of acting or another; one character may excel at comedic scenes, while another may perform best in tragedies. Nevertheless, it would be a solid measure of the empathic prowess of a character if it were able to elicit strong performances from human actors in a variety of different dramatic scenes.

5. FUTURE WORK

The key element of future work in this study is to complete the initial experiment described in section 3.2. The results from this study will form the baseline for further experiments in this domain. In addition, there are a number of other areas of potential future work that have been inspired by this methodology.

The first area in which we would extend the study would be to use the experimental design to evaluate the human actors rather than the heterogeneous acting partners. For example, it may be the case that actors of different skill levels are able to sustain acting quality to different extents over repeated performances. Also, highly trained individuals may be able to give consistent performances with a variety of "acting partners," while the performances of less well-trained actors are solid with human counterparts but fall off with less engaging stimuli. By examining the decay profile of an actor's performances over these two factors – different acting partners and number of repetitions – it may be possible to determine how far along in his or her training the actor is.

Second, we would like to help generate a set of baseline characters and scenes by which to allow other researchers to compare their characters. Offering several characters of different qualities could provide a metric by which to measure the relative abilities of different characters.

6. CONCLUSION

The core goal of this research is to offer a scientific method for comparing the empathic ability of two or more entities. The craft of acting makes significant use of empathy, as actors feed off each other to create higher quality performances. This paper proposes a methodology based in dramatic acting that allows researchers to compare the empathic value of several different characters.

The essential steps of the experiment involve collecting a group of characters, causing human actors to interact with each one in turn, videotaping these interactions, showing the videotaped interactions to skilled acting instructors and asking the acting instructors to evaluate each performance. Those characters who produce the highest quality performances in the actors who interact with them could be considered to have the greatest empathic ability.

The broader goal of this work is to contribute to the process of developing compelling and empathic synthetic characters by offering a concrete way of measuring very different characters against each other.

7. ACKNOWLEDGMENTS

The author would like to thank Joe Rojas and the other members of the Social Code Group at UCI. Discussions with Robert Cohen, Eli Simon, Phil Thompson and Richard Brestoff helped shape the form of the experiment. This project was made possible through the support of the ACE program and the Drama and Informatics Departments at UCI.

8. REFERENCES

[1] S. Winston, 2004, Personal Communication (Email), 2/3/04.

[2] Y. Yamanouchi, "Seamless image composition using actual and virtual studio sets…Advanced Virtual Studio," in *NHK Research & Development*, 2002.

[3] E. Stotland, K. Mathews, S. Sherman, R. Hansson, and B. Richardson, *Empathy, fantasy, and helping*. Beverly Hills: Sage Publications, 1978.

[4] R. Cohen, *Theater, Brief Version*. Boston: McGraw Hill, 2003.

[5] J. Gratch, J. Rickel, E. André, N. Badler, J. Cassell, and E. Petajan, "Creating Interactive Virtual Humans: Some Assembly Required," *IEEE Intelligent Systems*, 2002.

[6] K. Perlin and A. Goldberg, "Improv: A System for Scripting Interactive Actors in Virtual Worlds," presented at Proceedings of ACM SIGGRAPH 96, 1996.

[7] D. Isla, R. Burke, M. Downie, and B. Blumberg, "A Layered Brain Architecture for Synthetic Creatures," presented at Proceedings of the International Joint Conferences on Artifical Intelligence (IJCAI), Seattle, WA, 2001.

[8] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson, "Integrated learning for interactive synthetic characters," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*: ACM Press, 2002, pp. 417--426.

[9] T. Bickmore, "Relational Agents: Effecting Change through Human-Computer Relationships," in *Media Arts & Sciences*. Cambridge, MA: Massachusetts Institute of Technology., 2003.

[10] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan, "Embodiment in Conversational Interfaces: Rea," presented at Proceedings of the CHI'99 Conference, Pittsburgh, PA, 1999.

[11] J. Bates, A. Loyall, and W. Reilly, "An Architecture for Action, Emotion, and Social Behavior," presented at Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World, 1992.

[12] W. S. N. Reilly, "Believable Social and Emotional Agents," in *School of Computer Science*. Pittsburgh, PA: Carnegie Mellon University, 1996.

[13] H. Schaub, C. Zoll, and R. Aylett, "Modelling Empathy: The EU-project VICTEC (Virtual Information and Communication Technology with Empathic Characters)," presented at Fifth International Conference on Cognitive Modeling, Bamberg, Germany, 2003.

[14] H. Prendinger, J. Mori, S. Saeyor, K. Mori, N. Okazaki, Y. Juli, S. Mayer, H. Dohi, and M. Ishizuka, "Scripting and Evaluating Affective Interactions with Embodied Conversational Agents," *Künstliche Intelligenz (KI) Zeitschrift*, vol. 1, pp. 4-10, 2004.

[15] J. Velasquez, "When Robots Weep: Emotional Memories and Decision-Making," *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 1998.

[16] C. Breazeal, "Sociable Machines: Expressive Social Exchange Between Robot and Human," in *Artificial Intelligence Laboratory*. Cambridge, MA: MIT, 2000.

[17] M. Cavazza, F. Charles, and S. J. Mead, "Interacting with Virtual Characters in interactive storytelling," presented at Proceedings of the first international joint conference on Autonomous agents and multiagent systems, 2002.

[18] A. Paiva, I. Machado, and R. Prada, "Heroes, villians, magicians, …: dramatis personae in a virtual story creation environment," in *Proceedings of the 6th international conference on Intelligent user interfaces*: ACM Press, 2001, pp. 129--136.

[19] D. Thalmann, H. Noser, and Z. Huang, "Autonomous Virtual Actors Based on Virtual Sensors," in *Creating Personalities*, R. Trappl, and Petta, P., Ed. Heidelberg: Springer Verlag, 1997.

[20] S. Strassman, "Semi-autonomous animated actors," presented at Proceedings of the 12th National Conference on Artificial Intelligence, 1994.

[21] N. Magnenat-Thalmann, P. Kalra, and M. Escher, "Communicating with virtual characters," presented at WSCG '98, 1998.

[22] J. K. Hodgins and N. S. Pollard, "Adapting Simulated Behaviors For New Characters," presented at Proceedings of ACM SIGGRAPH 97, 1997.

[23] B. Hayes-Roth, E. Sincoff, L. Brownston, R. Huard, and B. Lent, "Directed Improvisation with Animated Puppets," presented at Proceedings of CHI '95, 1995.

[24] M. Assanie, "Directable Synthetic Characters," presented at AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, 2002.

[25] B. Reeves and C. Nass, *The Media Equation*. Cambridge: Cambridge University Press, 1996.

[26] B. Friedman, P. H. Kahn, Jr., and J. Hagman, "Hardware companions?: what online AIBO discussion forums reveal about the human-robotic relationship," in *Proceedings of the conference on Human factors in computing systems*: ACM Press, 2003, pp. 273--280.

[27] W. Tomlinson, "Synthetic Social Relationships for Computational Entities," in *Media Arts & Sciences*. Cambridge, MA: Massachusetts Institute of Technology, 2002, pp. 202.

[28] S. Turkle, "Sociable Technologies: Enhancing Human Performance When the Computer is not a Tool but a Companion," in *Converging Technologies for Improving Human Performance*, M. Roco and W. S. Bainbridge, Eds. Dordrecht: Kluwer Academic Publishers, 2003.

[29] D. Diderot, "The Paradox of Acting. (originally published 1830)," in *Masks or Faces*, W. Archer, Ed. New York: Hill and Wang, 1957.

[30] E. Simon, *Masking Unmasked: Four Approaches to Basic Acting*. New York: Palgrave Macmillan, 2003.

[31] R. Brestoff, *The Camera Smart Actor*. Lyme, New Hampshire: Smith & Kraus, 1994.

[32] P. Ekman, "An Argument for Basic Emotions," in *Basic Emotions*, N. Stein, and Oatley, K., Ed. Hove, UK: Lawrence Erlbaum, 1992, pp. 169-200.

[33] W. Shakespeare, *Macbeth*: http://the-tech.mit.edu/Shakespeare/macbeth/macbeth.5.5.html, 1623.