HRI: Four Lessons from Acting Method

GUY HOFFMAN
MIT Media Lab

Abstract

This informal paper discusses four principles found throughout much of the acting method literature that may be of value to Human-Robot Interaction (HRI) design: (1) psycho-physical unity, (2) mutual responsiveness, (3) objective and inner monologue, and (4) the taking into account of circumstances. This work is founded on the notion that both acting and HRI design are concerned with physically mimicking natural behavior in a collaborative setting.

HRI DESIGN AS ACTING

Research in interactive robots—those designed to work in close collaboration with humans—is concerned, among others, with the creation of natural behavior for the robotic character. This is particularly true for entertainment robots that do not serve a functional purpose beyond the interaction or performance they are expected to display. This paper argues that there are interesting parallels between the challenges actors take on and the ones that designers of HRI[1] robots need to tackle, and therefore lessons to be learned from acting methods for HRI development.

This work should be considered as food for thought, rather than a thorough review of acting literature or a fleshed out roadmap for HRI design. Its statements are best read as inspiration rather than as recommendation.

That said, why even look towards acting when thinking about robot design?

Relating the two fields stems from the observation that actors are trained to be highly tuned to the technical physicality of behavior for various actions and attitudes. An actor's preparation of a role includes a systematic investigation of what gesture, body pose or physical action best describes the internal drive and objective of their character in different contexts. Good actors pay attention to conventions of nonverbal communication and often need to take on the difficult task of portraying complex emotions without words.

[1] Human-Robot Interaction

It seems to me that these activities are similar to the ones one engages in when designing robotic behavior for human-centric robots. There, too, one is concerned with the readability of expressive behavior, with the breakdown and technical analysis of what it takes to convey a complex inner development, and with the translation of all that into what finally results in a series of motion commands.

To state it more boldly: both human-centric interactive robot designer and actor are in the business of injecting life into a lifeless object, be it a hierarchy of joints and motors for one, or an arc of dry dialog lines and stage directions for the other. Both robot designer and actor have to analytically break down the complex emergent constellation called "behavior" and reconstruct it in an inherently fake, but ultimately meaningful—or at least believable—way.

As teachers of acting method have been working to establish principles aimed to help actors prepare an analysis of human behavior with the proclaimed goal of artificially re-creating it in an interactive session, their methods may hold valuable lessons for the progress of HRI within and beyond the realm of entertainment robots. This paper explores these lessons.

There are, of course, significant differences between human-centric robot behavior and acting. Most notably, as Lee Strasberg said, "the actor need not imitate a human being. The actor is himself a human being and can create out of himself."[2] This obviously does not hold true for a robot, a fact that can be viewed both a curse and a blessing. On the one hand, much of acting technique draws on the personal experiences of the actor, experience that no robot can bring to the table. On the other hand, a significant portion of an actor's most challenging training is purposed to enable her to "let go" of ingrained experience, of her own personality, and allow the fictitious character to enter into that void. In that respect, robots are fortunate to not have a real ego to battle in an attempt to make room for their alter—or stage—ego.

This important difference could lead to the conclusion that there can be no robotic "acting" without there first being a robot personality. This also reflects a popular concern in Artificial Intelligence that predicts that artificial agents will never be intelligent until they are able to accumulate experience and memories over a prolonged period of time.

I agree that in order to seriously apply Stanislavskian methodology to robots, vast personal experience must first be acquired. But does that mean that the connection between robotics and acting is rendered impossible un-

[2] In [Cole and Chinoy, 1970], p.626

til robots are able to gather enough "life" experience to truthfully perform the "as if" projection that actors are required? Must we wait until robots experience as much as we do before the bridge between acting and robotics can be meaningful? I do not share this pessimistic outlook.

What, then, can acting theory teach us that can be applied towards human-interactive robots today? A survey of acting method training and theory literature, written primarily in the last few decades, demarcates several threads that reoccur in almost every text and technique, no matter how different from each other (and how much each method claims to revolutionize its predecessor's approach).

The persistence of these threads could indicate that they hold an essence of that life-inducing mechanism that good actors master, but thus far no robotic designer comes close to.

FOUR SCARLET THREADS

Four themes seem to repeat throughout most of the cited work on acting method. They will be referred to herein as (1) *psycho-physical unity*; (2) *mutual responsiveness*; (3) *objective and inner monologue*; and (4) accounting for the *context and circumstances of an action*.

In the following paragraphs I will briefly show how these themes have been addressed by various acting teachers and theorists, and why I think they may provide a possible lesson for HRI designers.

(1) Psycho-physical Unity

In his book *Games for Actors and Non-actors*, Augusto Boal states quite elegantly:

> "[...] the human being is a unity, an indivisible whole. Scientists have demonstrated that one's physical and psychic apparatuses are completely inseperable. Stanislavski's work on physical actions also tends to the same conclusion, i.e. that ideas, emotions and sensations are all indissolubly interwoven. A bodily movement 'is' a thought and a thought expresses itself in corporeal form."[3]

Generally hailed as a recent scientific discovery, mounting evidence shows not only the effects of our mental processes and psychological state

[3][Boal, 2002], p.61

on our corporal functions but—maybe more surprisingly—the strong influence our physical behavior exerts on our mental state. However, it is safe to say that the notion of a psycho-physical link existed for many centuries, if only as an intuition ("*mens sana in corpore sano*").

Acting theorists have acknowledged that behavior is inherently physically grounded. As early as the late 19th century, acting system reformer François DelSarte, notes the feedback from body to mind:

> "A perfect reproduction of the outer manifestation of some passion, the giving of the outer sign, will cause a reflex feeling within."[4]

This interrelationship is only gaining momentum since. Viola Spolin muses "through physical relationship all life springs"[5], Robert Lewis stresses the physical and sensory side of *affective memory*[6], Acting teacher Sonia Moore reminds us that "it is a fact that in life the whole complex inner world of a human being, every inner experience, is always expressed physically"[7], and stresses the "how we do it"—the physical aspects of the emotional motivation—as a central part of her Stanislavskian method. Ruth Maleczech uses images, corporal representations, to elicit her performance[8], as does Boal in his "Image Theater". Broadhurst writes: "[...] technology's most important contribution to art is the enhancement and reconfiguration of an aesthetic creative potential which consists of interacting with and reacting to a physical body, not an abandonment of that body"[9]. And the list goes on. To conclude with Strasberg's more poetically leaning words: "The actor has no piano. In the actor, pianist and piano are the same"[10].

Following this strong indication, HRI researchers should feel confident for exploring the physically grounded intelligence of robots rather than abstract computational processes. The fact that robots actually *have* bodies is a good starting point, and much of the Aritficial Intelligence community seems to have finally woken up to the message that intelligence is in-

[4][Stebbins, 1887], p.63
[5][Spolin, 1999], p.16
[6]In [Cole and Chinoy, 1970]
[7][Moore, 1968], p.94
[8]In [Sonenberg, 1996]
[9][Broadhurst, 2004]
[10]In [Cole and Chinoy, 1970]

herently embodied. When it comes to human-robot interaction, thinking without moving might well turn out to be a mostly hopeless endeavor.

However, in HRI robots designed today cognitive models are still distinctly separate from their physicality. While admittedly immersed in sensory experience, not much attention has been paid to the physical representation of cognition. The "behavior systems" of robots are mostly action selection mechanisms driven by sensory input, and in most HRI implementations the virtual agent driving the robot is an extension of an ideal logical process that communicates with the physical results of its decisions as input and output streams to and from a distant module.

One recent trend in cognitive psychology (see for example the work of Larry Barsalou[11]) supports what most acting teachers seem to know well: the relationship between mind, decision and corporal function is not one of two distinct systems drawing from each other in a remote-controlled way, but are two aspects of the same; two sides of a "unity".

The one field of robotics where a similar message seems to be most prominently incorporated is that of close feedback motion control, in which intelligence is employed towards simple, bounded, physical activity—such as balancing a pole or stabilizing a walk. These applications, however, are contained to the most rudimentary activities, and do not scale towards the physical aspects of higher level cognition, which probably should be viewed in the same manner.

While the basic message is clear, it is not completely obvious how the robot's physical motion could be incorporated into a better model of robot cognition, but—regardless—I strongly feel that it must.

Robot design should consider tearing down the implicit barrier between the "motor system" and the "behavior system" and think afresh about a combined architecture where both are sides of the same behavior. Motion should not only influence "thinking" (as it sometimes, but rarely, does), but should *be* the decision process.

A conceivable way forward is to constrain artificial cognitive process to physical embodiments of that process, not allowing any "thinking" that is not physical in nature, simulating more closely our own mento-corporal dependency.

[11][Barsalou et al., 2003]

(2) Mutual responsiveness

New York City acting guru Sanford Meisner is generally associated with the stressing of responsiveness in acting. His rule, embodied among others in the famous "repetition exercise", states

> "Don't do anything unless something happens to make you do it. What do you do doesn't depend on you, it depends on the other fellow." [12]

In another place, he commands that "acting isn't chatter, it's responding truthfully to the other person".[13] This rule, in Meisner's method, is the key to truthful reaction, and meaningful behavior within the on-stage collaboration between two actors.

Meisner is by no means the only acting theorist to find that what is really occurring in a scene is not happening within any of the actor's heads, or even in their behaviors, but in the space between the two actors. Spolin calls this *communication*:

> "[T]he techniques of the theater are the techniques of communicating. The actuality of the communication is far more important than the method used."[14]

Similarly, Maleczech speaks of *repercussion*:

> "The other actors are, for me, like the bumpers in a pinball machine. I shoot my pinball, my image, and it goes *tch, tch, tch*, bouncing off those bumpers, each hit having its repercussion. Often the next image will come directly from the response of the other actor."[15]

Sonia Moore reminds us that "as in life, you must evaluate the other person and expected results,"[16] as well as "you must coordinate your behavior. [...] Ensemble work means continuous inner and external reaction to each other."[17]

[12] [Meisner and Longwell, 1987], p.34
[13] Ibid, p.41
[14] [Spolin, 1999], p.14
[15] In [Sonenberg, 1996], p.91
[16] [Moore, 1968], p.104
[17] Ibid, p.106

Of all four, this should perhaps be the most optimistic of fronts in applying acting knowledge to interactive robots. After all, robots, due to their impoverished internal lives and limited experience, tend to be inherently reactive machines. From their very roots, interactive performance robots (as early as James Seawright's Searcher (1966) and Watcher (1966)) "moved and emitted different sound patterns in response to movements and light changes going on around them".[18]. So it seems that building a robot that can engage in *some* version of the repetition game might not to be an impossible feat.[19]

However, pure mechanical repetition can hardly be considered truly responsive. The missing link in order for repetitive agents to be useful partners in Meisner's repetition game, is solving the undoubtedly harder second part of the game, which is to *know when to change the repetitive behavior*. For mutual responsiveness is not merely responding in a predictive way, but knowing when to break the mirroring and as a result—switching one's action. It is this failure to break away from the predictive pattern, that causes realistic mutual responsiveness to seem, for the most part, broken in HRI robots.

We should expect to make important strides in the field of robot responsiveness, since of all the acting threads discussed in this paper, this one seems to be most suited for automated entities. To do so we must solve the difficult problem of choosing between simple reactive behavior and breaking away from simple reaction. Detecting the right time and manner to do that is a hard perceptual challenge, but may well hold the key to solving the mutual responsiveness problem.

(3) Objective and Inner Monologue

Perhaps the most significant contribution to modern acting that has been brought on by the Stanislavski system is the elimination of so-called "representational acting". This method of formalizing a range of emotions and associating them with certain bodily and facial actions is often attributed to the *DelSarte system of expression*, a methodology that has unsurprisingly caught the attention of some robot designers.[20]

[18] [Dixon, 2004]

[19] It could be argued that the early interactive computing project *ELIZA* was a version of a repetition-game agent.

[20] [Bruce et al., 2001]

> "The actor in the French school [...] asks himself 'What must I *choose to do*?' "[21]

Teachers of the Stanislavski method stress a decidedly different approach, in which the character's motives, objectives, and intentions weigh stronger than the actual line they deliver, allowing the actor to rise beyond simple representation.

Sonia Moore quotes Eugene Vakhtangov saying "A unit in a role or a scene is a step in moving the through-line of actions toward the goal,"[22] as opposed to the unit being a single line of dialog or a stage direction.

Her method also encourages the actors to not memorize the lines, but instead focus on analyzing a scene in terms of moving powers, objectives, obstacles, and intentions, leading to choosing actions, creating images and "what if" behaviors out of *affective memory* to truly understand a scene. The scene is then improvised with disregard to the text, and finally re-implemented out of that improvisation on the actual lines of the scene.

Other teachers share this view. Michael Chekhov speaks of Psychological Gestures that draw on the character's "definite desire" in a scene[23], and Boal, too, stresses that any particular action results from the character's desires, will, and needs.

This then leads to the concept of the *inner monologue*, which is heavily emphasized by Moore's method (and referred to by Meisner as well).

> "[Your inner monologue] is more important than memorizing your lines [...]. The right inner monologue will bring you to your lines, and you may have entirely different intonations." [24]

The actor's inner monologue is based on the analysis of the character and must carry on the whole time the actor is on stage, whether she says something or not. This inner monologue should usually be laid out in detail while preparing for a role, and lends the actor credibility of an internal process while they're on stage, leading up to the lines and thus preventing the lines to be uttered in a void.

[21] In [Cole and Chinoy, 1970], p.626; emphasis mine.
[22] [Moore, 1968], p.142
[23] In [Cole and Chinoy, 1970], p.519
[24] [Moore, 1968], p.106

From my own experience with designing behavior for an expressive HRI robot, as well as from my observation of other HRI robots it is evident that an underlying intention structure is immensely conducive to a reasonably *behaving* robot, intelligence aside. This is particularly true if we attempt to create life-like behavior.

Cracks showing in a robot's lifelike behavior are most often due to the surfacing of too simple, or worse—random behavior in the action selection system. The robot seems to be behaving unintelligently whenever it displays a lack of clear and readable objectives, desires and intentions.

In our research group we have also found that as little as the most simple combination of an underlying desire and affective state modulating the otherwise straightforward functional behavior of the robot significantly changes the way people interact with the robot, attributing a much higher degree of ”understanding” and ”realism” to the inanimate object. Our implemented combination of both a reactive system and an internal state system allows the robot to operate in a much more complex behavior space. This is in contrast to the usual discrete-state graph that is associated with interactive robots.

Obviously, I do not believe that robots will have as rich an intentional mechanism as we do, but there might even be value in applying a simplistic objective-based system in order to achieve more complex behavior patterns. If we are attempting to create interactive robotic actors, the notion of objectives and desires must exist at the base of the robot's behavior, even if they are based merely on a very rudimentary motivational system.

More specifically, I feel that the idea of an “inner monologue” is a powerful one that can, and should, be transferred to the realm of Human-Robot Interaction. Just as an actor should be constantly conducting an inner monologue to achieve continuity and realism in his behavior, and to make his externally evident actions emotionally based, it makes sense for an HRI-centered robot to continuously progress internal processes, instead of “waiting” for the human and then selecting the correct response.

We might find that endowing an HRI robot with “inner monologue” will do the same trick it does for actors, i.e. result in a much more natural and continuous interaction, as the robot picks up impulses (in the Meisner meaning of that word[25]) issued by the human before it actually responds to those impulses. This could be a significant key in avoiding the currently

[25] [Meisner and Longwell, 1987], p.72

prevalent command-and-response behavior robots usually display.

With inner monologue, sparse actions can become tips of icebergs of internal processing, rather than a disconnected array of floating islands of behavior.

Inner monologue must, of course, play out externally—cueing the human counterpart as it evolves, leading up to and filling up blank spaces between actions, and adding an intriguing layer of nonverbal communication to human-robot interaction.

(4) Context / Circumstances

Compared to the other three, this thread is not as prominently discussed in the acting literature. It is, however, a cornerstone of the Stanislavski method, which teaches that the same action and the same dialog can have a substantially different content given different circumstances. To illustrate, improvising on the action of "hiding", Sonia Moore explains:

> "Do not think of how you will perform it before you have a clear picture of the circumstances in your mind. Different circumstances will make you hide in very different ways. It must be obvious to you that hiding from children in a game is different than hiding from a gangster who is following you. [...] [Only] after you know when, where, why, and so on, you must think of how—" [26]

The notion of the circumstances of a certain event is at best in its infancy in Artificial Intelligence, and exists only very primitively in most robotic applications. While finding its way into the Natural Language Processing literature, the concept of choosing different actions, and different modulations to these actions based on different contexts, is absent from most interactive robot design.

How can we use this idea in for HRI robots? Some interactive character research has been experimenting with "adverbs" to modulate blended character animation based on internal context, but the road to fully take into account the circumstances of a situation is mostly unpaved.

State-Action decision models are for the most part the closest approximation to context-aware behavior, with their success heavily dependent on a useful modeling of the state space. Moreover, even when state space modeling is done well, there is no adequate model to direct an artificial agent's

[26] [Moore, 1968], p.38

attention as to what aspect of the state should be taken into account as the *relevant context* for a particular action. This usually leaves the agent having access to either too much information or too little thereof (and often both). It would be safe to say that our model of how to take into account context and circumstances in A.I. decision making leaves much to be desired for.

While this paper is far from claiming to offer a solution for this hard and much discussed problem, I propose to take note of acting methodology's way of coping with this issue. Many improvisational exercises focus on training how specific circumstances change the same theatrical action. Perhaps a close observation of such techniques might clue us in on how states could be modeled to more accurately represent context and circumstances.

SUMMARY

Actors have to solve some problems that are surprisingly related to the ones HRI designer are struggling with. This paper identifies four persistent threads appearing throughout most acting theory and practice literature. While not providing resolute answers, they point us in interesting directions. Moreover, their pervasiveness may indicate that they hold central ideas that make acting possible, rendering them candidates for serious consideration by HRI designers.

First, I reviewed the notion of the *psycho-physical unity*. In today's interactive robots, behavior systems and motor systems are separate, if communicating, units. Acting method suggests to unify the two to be aspects of the same process. Second, *mutual responsiveness* has been shown to be a cornerstone of—among others—Meisnerian method. Focusing on the *between* instead of the *within* could prove to be a promising direction for rethinking HRI design. Third, Stanislavkian teaching suggests analyzing action in terms of *objectives and inner monologue*. Taking the 'tip of the iceberg' approach to Human-Robot Interaction might make human-centric robots behave in a more principled, continuous, and natural fashion. Finally, a fresh look at modeling *context and circumstances* is direly needed in A.I. decision making theory. Actors' preparatory exercises aimed at this aspect of character analysis could hold valuable clues for that endeavor.

REFERENCES

[Barsalou et al., 2003] Barsalou, L. W., Niedenthal, P. M., Barbey, A., and Ruppert, J. (2003). Social embodiment. *The Psychology of Learning and*

Motivation, 43:43–92.

[Boal, 2002] Boal, A. (2002). *Games for Actors and Non-Actors*. Routledge, 2nd edition.

[Broadhurst, 2004] Broadhurst, S. (2004). The Jeremiah Project: Interaction, Reaction, and Performance. *The Drama Review*, 48(4).

[Bruce et al., 2001] Bruce, A., Nourbakhsh, I., and Simmons, R. (2001). The role of expressivness and attention in human-robot interaction. In *Proceedings of the 2001 AAAI Fall Symposium*, Cape Cod, MA, USA. AAAI Press.

[Cole and Chinoy, 1970] Cole, T. and Chinoy, H. K. (1970). *Actors on Acting : The Theories, Techniques, and Practices of the World's Great Actors, Told in Their Own Words*. Crown, 3th edition.

[Dixon, 2004] Dixon, S. (2004). Metal Performance: Humanizing Robots, Returning to Nature, and Camping About. *The Drama Review*, 48(4).

[Meisner and Longwell, 1987] Meisner, S. and Longwell, D. (1987). *Sanford Meisner on Acting*. Vintage, 1st edition.

[Moore, 1968] Moore, S. (1968). *Training an Actor: The Stanislavski System in Class*. Viking Press, New York, NY.

[Sonenberg, 1996] Sonenberg, J. (1996). *The Actor Speaks : Twenty-four Actors Talk About Process and Technique*. Three Rivers Press, 1st edition. Selected chapters.

[Spolin, 1999] Spolin, V. (1999). *Improvisation for the Theater*. Northwestern University Press, Evanston, IL, USA, 3rd edition.

[Stebbins, 1887] Stebbins, G. (1887). *Delsarte System of Expression*. Edgar S. Werner, New York, NY, 2nd edition.