Dramatic Expression in Opera, and Its Implications for Conversational Agents

W. Lewis Johnson
Director, Center for Advanced Research in Technology for Education (CARTE)
USC / Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292 USA
Johnson@isi.edu

Introduction

It is commonly agreed among embodied conversational agent (ECA) researchers that ECA behavior should be based upon principles of human face-to-face communication (Cassell et al., 2000; Traum & Rickel, 2002). It is less commonly acknowledged that principles of human *acting* can inform the design of ECA behavior, particularly in making behavior engaging and understandable. Character animators, in contrast, understand clearly the relationship between character behavior and acting (Porter, 1997), and have articulated principles such as exaggeration and staging that are based in part on observations of actors (Thomas & Johnston, 1981; Lasseter, 1987; Maestri, 1999). However, we cannot expect to capture principles of dramatic portrayal in ECAs simply by copying the techniques of animators. ECAs are being developed for a applications with a variety of media characteristics; we therefore need to draw lessons from a range of dramatic media, including those involving live action. Some ECA developers try to incorporate dramatic aspects by collecting motion capture data from actors (Churchill et al., 2000). This approach relies upon the actor's expressive skills to achieve the desired dramatic effect. Unfortunately there is no assurance that motion capture data will appear equally expressive and appropriate when transferred to different media and different dramatic contexts.

This article considers dramatic portrayal from a personal perspective: that of a practicing opera singer. Through examination of the process of preparing and performing an operatic role, I will attempt to draw lessons that may be of value to the design of conversational agents, and discuss how those lessons apply to specific examples of conversational agents. Lessons learned here are particularly applicable to ECA applications dealing with emotional or stressful subjects, those that involve long-term interaction with agents, and those that seek to engage the user deeply in the subject matter. To those of you who are not well acquainted with opera as a dramatic form and doubt its relevance to ECAs, I suggest that you try to suspend your disbelief, and read on.

Case Study: *Susannah*

I recently completed a stint as an opera singer, performing the role of Olin Blitch in Carlisle Floyd's opera *Susannah*. The production was mounted by Ventura College, and performed by a mix of professionals and amateurs from the Ventura, Santa Barbara, and Los Angeles areas. The story of *Susannah*, based upon the biblical story of Susannah and the Elders, is set in the Bible Belt of Tennessee. The elders of the local church chance upon Susannah bathing naked in a nearby creek, and accuse her of being a sinful woman. They inform an itinerant preacher, Olin Blitch, who has just come into town to lead a revival meeting, and Blitch resolves to try to convince Susannah to repent. Susannah refuses, because she believes that she has done nothing wrong. Blitch takes her refusal as evidence that she is beyond redemption, and proceeds to seduce her (Figure 1). Only then, when he discovers that she is a virgin, does he realize that Susannah was unjustly accused, and that he is the one who now faces eternal damnation.

Floyd termed his work a "musical drama," and it is clear that in this work dramatic goals are paramount. The use of multiple modalities, including music, enhances the dramatic portrayal.

On the other hand, the performer's need to convey intent at a distance from the stage, and the need to be heard over a powerful orchestra, make it necessary to intensify expression at times. But because the dramatic expression is intense, the mechanisms involved in its portrayal are relatively easy to discern and study.

Figure 1. Blitch and Susannah

Points of Comparison: Carmen's Bright IDEAS and MRE

This article will make frequent references to two particular ECA systems that I have been involved in developing: *Carmen's Bright IDEAS* (CBI) and Mission Rehearsal Exercise (MRE). I will describe these systems briefly here; for more detail please see the cited references.

Figure 2. Points of comparison: Carmen's Bright IDEAS and Mission Rehearsal Exercise

Carmen's Bright IDEAS is an interactive pedagogical drama designed to teach mothers of pediatric cancer patients to cope better with their problems (Marsella et al., 2003a; Marsella et al., 2003b). It dramatizes the problems of Carmen, a fictional mother of a child with cancer, and shows Carmen discussing her problems with a counselor, Gina. Agent technology is used to

determine Carmen's and Gina's actions, and neither character's behavior is scripted ahead of time. *CBI* was developed with dramatic concepts in mind; the story was developed initially as a linear script by a professional scriptwriter, and then extended into a library of possible actions for each character. Character gestures and facial expressions were designed to ensure that the intent of the characters is clear to the viewer (Figure 2).

Mission Rehearsal Exercise (MRE) is designed to train military leadership skills in stressful situations (Swartout et al., 2001). MRE was developed by USC's Institute for Creative Technologies, in collaboration with CARTE and USC's Integrated Media Systems Center. MRE places trainees in a simulated peacekeeping situation where they must interact with simulated platoon members and make decisions. The action takes place on a large floor-to-ceiling panoramic display that gives an illusion of presence in the virtual scene. The scenario consists of three main scenes. The first scene gives a first-person view of driving into town. The second scene takes place in the town where a traffic accident has occurred between a military vehicle and a civilian car (Figure 2). The climax of the action takes place here. Finally there is a third scene consisting of a fictional television news report summarizing the outcome of the scenario.

Some Observations and Lessons

Dramatic structure

Large-scale works such as operas have a dramatic structure that helps to promote audience engagement. Individual scenes such as the revival meeting in *Susannah* have a progressive build-up of dramatic intensity. Likewise the sequence of scene leads to the overall climax of the work. Freytag suggested a canonical form for the dramatic structure of such works, called "Freytag's triangle", which consists of rising action leading to a climax, followed by falling action leading toward the conclusion (Freytag, 1898, cited by Laurel, 1991). Yet simple structures such as Freytag's triangle do not however capture the full structural complexity of operatic works. Instead, action develops over a series of intermediate climaxes, often followed by contemplative scenes in which the characters reflect on what just happened and decide what to do next.

Dramatic structure was also a factor in the design of *CBI* and MRE. They employ a three-scene structure, in which the main scene provides the climax. A major challenge in these systems is ensuring that each session exhibits proper dramatic structure regardless of what actions the autonomous characters and the user take. For example, Gina supports the dramatic structure of CBI by guiding Carmen through the problem solving steps, when she feels that Carmen is ready to continue. If Carmen, under the influence of the human learner, veers away from the intended dramatic structure, e.g., by losing confidence and refusing to develop options, Gina tries to motivate Carmen to get back on track.

One thing that is needed in titles such as CBI and MRE is a dramatic structure that links multiple sessions. We want each session to have a dramatic resolution, and yet motivate the learner to continue to work through multiple training sessions. This is a common challenge for agent-based applications that interact with users over multiple sessions. Serialized dramatic forms might serve as useful models here. But it may also be possible to adapt the operatic technique of interspersing intermediate climaxes with more reflective scenes, to examine what has just happened and prepare the user to continue the story in subsequent sessions.

Character development

In addition to the overall dramatic arc of the story, operas typically also incorporate development arcs for each main character. In *Susannah* each character arc starts with a clear expository scene, in which the character expresses thoughts and intentions, so that the audience understands the

motivations for his or her subsequent actions. As the story unfolds more facets of the characters' personalities may be revealed, as they react to new events. We see Blitch go through a series of changes, from upright preacher to seducer to repentant sinner. Floyd has written arias for Blitch at each major change, in order to make the changes clearer to the audience. This is important because, as dramatic theorists since Aristotle have noted, the character's actions should follow causally from the character's traits, and that the character's traits should be consistent throughout (Telford, 1961). So if a character changes over time there must be a cause for this change, from the audience's perspective; these changes may be a consequence of significant plot events, or may reflect additional character traits that have not yet been revealed. When the character's traits are pulling him in conflicting directions, e.g., when Blitch decides whether or not to seduce Susannah, the audience must see that conflict, so that characters actions do not appear arbitrary.

To build character arcs into conversational agents, we need to keep the agent's character traits consistent, while providing mechanisms for development and change. There have been significant advances recently in defining agent character traits; Rist et al. (2003) have developed a toolkit for specifying personality traits in accordance with Dignum's Big 5 model. Gratch's Émile system models emotions using a plan-based model of emotional appraisal, and shows how small biases in emotional appraisal and response can lead to large systematic differences in character behavior (Marsella et al., 2003a). These mechanisms have been integrated into MRE, yielding agents that can respond in different ways to external events, based upon personality parameters (Marsella & Gratch, 2003). Nevertheless, these models draw a strict dichotomy between agent moods, which are ephemeral, and personality traits, which are fixed. The middle ground, of evolvable yet consistent character traits, needs further development in ECAs.

One potentially valuable mechanism comes from research on the psychology of motivation, which has identified a number of motivational factors that contribute to learning and achievement, such as confidence (Lepper & Henderlong, 2000) and fear of making mistakes (Linnenbrink & Pintrich, 2000). These factors are persistent, but are influenced by events.

Another possible technique is to model explicitly the "front" that characters present to others. Goffman (1959) has observed that people in social situations try to present themselves in a manner that is appropriate to that situation. They attempt to manage both the expressions that they *give*, i.e., the communicative acts aimed at specific people, and the expressions that they *give off*, i.e., actions that others treat as symptomatic of them, that help others to form an impression. A character like Blitch is very much involved in presenting fronts to people, for example when he arrives in town and tries to assume spiritual leadership of the town. When I played Blitch in this context I made a point of conveying to the townspeople a confident, charismatic, empathetic, commanding persona, through his posture, his hand and facial gestures, and his stance and interpersonal distance during conversation. Later in the drama I had Blitch drop that persona, both in his interactions with Susannah and in his confession to God. Finally at the end Blitch tries to reassume his preacher persona in dealing with the townspeople, but he can no longer do it convincingly because he knows in his heart that it is false. Thus a character arc develops through a progression of social stances combined with changes in beliefs and attitudes. In the process, the audience gets to see to some extent behind the fronts that the character presents, and develops a consistent understanding of the character.

Verbal expression

Although many of the details of operatic portrayal are written into the score, many other details are omitted, and are up to the conductor, the director, and the singer-actors to create. I will focus here on the verbal aspects of operatic portrayal (i.e., song and speech), emphasizing those aspects that are relevant to conversational agents; nonverbal aspects will be discussed in the next section.

Susannah employs a wide range of verbal delivery, including conversational speech, half-sung *Sprachstimme,* sung recitatives imitating conversational cadences, and arias.

An important aspect of any verbal expression in opera is its emotional content. Singer-actors have multiple means at their disposal for expressing emotion, including tempo, volume, pitch range, accents, phrase shape, vocal color, and even vocal gestures such as sighs and tremors. The dynamic markings in the score only provide a rough guide to these qualities, and omit important details. But even they indicate characteristics such as volume they do not indicate the underlying rationale for the dynamics. In order for the dynamics to be convincing a performer should infer or imagine the intent underlying the dynamic marking, and try to express the intent.

This expression of intent is not simply a matter of displaying emotion—this emotion must be *communicated to* somebody. Emotional displays arise in the process of communicating to other characters. The manner and intensity of the emotional displays depend upon whether the singer-actor is communicating to an individual or a group, and the degree of familiarity of the listeners. Emotions can sometimes be displayed deliberately, to make the communication more persuasive. The context and communicative goals of expression are important because they influence both the focus of emphasis and the intensity of delivery. For example, when Blitch says to the church elders "Make restitution *now*, brethren!" he is not simply expressing anger, but is angrily uttering a command to them. This causes the entire utterance to be delivered at high intensity, with particular emphasis on the word "now." Intensity is also sensitive to the dramatic structure of the scene; if dialog is leading up to a climax, expressive intensity may increase accordingly.

Dramatic verbal expression poses serious challenges for conversational agents. Speech synthesis techniques that offer expressive variability usually have low speech quality. The text-to-speech synthesizer developed for MRE, in contrast, has good expressive qualities and overall sound quality, while providing significant expressive variability. It is a concatenative unit selection synthesizer that combines multiple limited-domain synthesizers, each specialized to a particular class of communicative intent (informing vs. inquiring vs. commanding) (Johnson et al., 2002). This helps to ensure that each utterance conveys the most suitable basic category of intent. We have recently extended the synthesizer to generate appropriate boundary tones depending upon the dialog context, and to emphasize particular words.

Dramatic gesture

Operatic portrayal employs a variety of nonverbal gestures—hand gestures, facial expressions, head and body poses, and body movement. Gestures complement the voice, making intent clearer and more compelling, and they extend portrayal through silent periods, when other singers are singing, or during musical interludes.

Gestural portrayal must work within strict constraints. Temporal constraints are imposed by the musical score, as interpreted by the conductor. Spatial constraints come from the blocking of the scene, requiring action to take place at set points on the stage and movement to proceed from one point to another. The singer-actors must determine what actions to perform and how within these constraints. I will not discuss the issue of blocking design here, but note that it is a complex problem, both for operas and for ECAs, particularly in multi-character scenes.

A major question is what range of gestures to use—should they be based on natural expression or stylized in some fashion? Contemporary singer-actors usually base their gestures on natural face-to-face conversational gestures. The main difference from normal conversation is that actors make greater use of their full body in conveying emotions and attitudes. I used posture extensively in my portrayal of Blitch, to depict his progression from confident man of God (erect, chest thrust forward) to repentant sinner (stooping, slope-shouldered).

Gestures must be natural, fit the constraints of score and blocking, convey intent effectively to the audience, as well as be aesthetically pleasing. Some experimentation may be required during practice and rehearsal to come up with a series of gestures that works most effectively. The danger, as Stanislavski (1936) has noted, is that the gestures take the place of the intention that the gestures are meant to express; the actor "represents" the part, instead of "living" it. What is required, according to Stanislavski, is an integration of inner intention and outer expression. I submit that this integration is important for conversational agents as well. One advantage of this approach, as Stanislavski has noted, is that it facilitates improvisation. If an actor memorizes a particular sequence of gestures to perform, that makes it difficult to adapt the portrayal if the drama unfolds in a different way from what was anticipated. Unexpected events can happen on stage, even in performances of linear dramas. The unexpected is even more likely to occur in nonlinear, interactive experiences. Similarly if an ECA is simply playing prerecorded gestures, acquired through motion capture or other means, without a model of the underlying emotional state, then if the situation changes unexpectedly the gestures may no longer appear appropriate.

Give and take

When multiple players are on stage, as is usually the case in opera, special considerations arise. There are often multiple activities going on at once, which is confusing since a viewer can only focus on one activity at a time. It is important for actors to coordinate their activities to make the overall action on stage understandable and coherent.

One way to lend coherence to multi-player action is to *give focus*. If one player has the primary role in the current action, then the other characters should direct their attention to that character. This helps the audience to see where to direct their attention, and avoids extraneous action on stage that can distract the viewer. We utilized this technique in the opening scene of *Susannah*. As Blitch, I made my entrance quietly, sat down, and listened to one of the townspeople, Mrs. McLean, talk about Susannah, whom she believes is evil. At this point I was giving focus to Mrs. McLean. Then after this Mr. McLean saw me, noted that I was a stranger, and asked me what my name was. I then stood up and announced in my first aria, "I am the Reverend Olin Blitch…" At this point everyone on stage directed their focus toward me, some turning to look at me and listen, some moving downstage so that they can get a better view of me.

Giving focus does not simply involve staring at other cast members, however. Each player must have an intention at all times, and display to that intention. So if a player is focusing on another player and listening to what that player is saying, the first player should react to what the other player is saying, and display that reaction. Action on stage involves a continual give and take among the players, where action leads to reaction which entrains further action. When done right these actions and reactions combine into a continuous flow, which propels the drama forward.

In order for give and take to work most effectively, the two players must work together so that each action provides preparation for the reaction. One mechanism of achieving this is through eye contact. If one player speaks or sings a line that calls for a strong reaction from the other player, he or she often will establish strong eye contact with the other player. This helps the other player to prepare to react to the action, and helps make the focus of action clear to the audience.

CBI and MRE both illustrate how give and take could apply to embodied conversational agents. During the vehicle accident scene in MRE a number of characters are present, but it is hard to tell what the focus of the action is. In Figure 2, for example, the mother and the combat lifesaver are focused on the boy, and the sergeant is focused on the viewer. This may be appropriate when the trainee first comes to the scene, but as the sergeant and the trainee plan how to evacuate the child the focus should shift to the trainee and the sergeant. Part of what makes the situation difficult in MRE is that the injured child and the lieutenant are competing foci of attention. In order to avoid

an appearance of lack of focus, transitions in focus from one point to another is necessary over the course of the action, in reflection of changes in saliency over time.

Taking the audience's perspective into account

Finally, I will discuss some of the ways in which stage action in opera takes the audience's perspective into account. Theatrical performance offers little in the way of direct interaction between the players and the audience. In fact, such interaction is discouraged, because it tends to lead to bad acting, and because it is difficult to establish give and take with an audience. Nevertheless, stage action on stage is carried out so as to make it understandable to the audience, and many of the techniques described above facilitate this. The following are additional ways in which operatic performance takes the audience's perspective into account; these may have relevance to ECAs, particularly those where the point of view of the audience is fixed, or under the control of the user instead of the agents.

One basic requirement is that the action be visible to the audience. Players must work to keep their action visible, particularly in dialog with other characters. An exchange between two characters, Blitch, and Elder McLean, illustrates this. Blitch, standing upstage from Elder McLean, wants to ask McLean a question. McLean will not be able to answer from this position, since it would involve singing upstage. Therefore Blitch needs to combine asking the question with walking downstage, to a position level with or downstage from McLean, and time his asking of the question so that he does not end up singing upstage either.

One way to make action more visible is to adjust body orientation toward the audience. If two players are standing side by side and engaged in conversation, they are each likely to turn slightly outward, rather face each other straight on, and reserve straight-on orientation for points where particular emphasis is required. This works in part because proscenium provides a two-dimensional frame for the action, making distortions of orientation less noticeable. ECA's could use this technique, since computer displays also frame action, and are usually two-dimensional.

Players must also take into account the distance of the audience. Gestures that read well close up may not be noticeable to audience members sitting at a distance. This means that gestures tend to be more pronounced on stage than in face-to-face conversation. ECAs are rarely life size, and in the future may increasingly appear on handheld devices. The problem of making gestures read on a small screen is similar to the problem of making gestures read at a distance.

Conclusions

This article has discussed principles, techniques, and methods of dramatic portrayal in opera, and their application to the development of embodied conversational agents. Investigations such as this complement studies of natural human behavior, and offer insights as to how to make such behavior understandable and interesting when adapted for use by embodied conversational agents. However, one should use caution in applying such lessons. The unique characteristics of computer-based media are still being identified and explored. In any case, one must always be careful about applying principles blindly to any artistic form. Such principles are post-hoc analysis of the intuitive skill of great artists; this was as true in Aristotle's day as it is today. We should not let structural principles stand in the way of injecting creativity into the design of ECAs. Opera at its best possesses an element of magic that is difficult to describe, much less analytically reconstruct. We can only hope to achieve a similar result with conversational agents.

Acknowledgments

The author wishes to thank Kate LaBore, Stacy Marsella, and L.T. Pirolli for their helpful comments, and funding from the Army Research Office and the National Cancer Institute.

References

Bickmore, T. (2003). *Relational agents: Effecting change through human-computer relationships*. Ph.D. thesis, MIT.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000). Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill, (Eds.), *Embodied Conversational Agents*, 29-63. Cambridge, MA: MIT Press.

Chekhov, M. (1953). *To the actor: On the technique of acting*. New York: Harper and Row.

Churchill, E.F., Cook, L., Hodgson, P., Prevost, S., & Sullivan, J.W. (2000). "May I help you?": Designing embodied conversational agent allies. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill, (Eds.), *Embodied Conversational Agents*, 29-63. Cambridge, MA: MIT Press.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Johnson, W.L, Narayanan, S., Whitney, R., Das, R., Bulut, M., & LaBore, C. (2002). Limited domain synthesis of expressive military speech for animated characters. IEEE TTS Workshop.

Lasseter, J. (1987). Principles of traditional animation applied to computer animation. *Computer Graphics* 21(4), July 1987, 33-44. New York: ACM Press

Laurel, B. (1991). Computers as theater. Reading, MA: Addison-Wesley.

Lazarus, R.S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lepper, M.R. & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research in intrinsic versus extrinsic motivation. In C. Sansone & J.M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance*, 257-307. San Diego: Academic Press.

Linnenbrink, E.A. & Pintrich, P.R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In In C. Sansone & J.M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance,* 195-227. San Diego: Academic Press.

Maestri, G. (1999). Character Animation 2, Volume 1: Essential Techniques. Indianapolis, IN: New Riders Publishing.

Marsella, S. & Gratch, J. (2003). Modeling coping behavior in virtual humans: Don't worry, be happy. In Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems. New York: ACM Press.

Marsella, S., Gratch, J., & Rickel, J. (2003a). Expressive behaviors in virtual worlds. In press.

Marsella, S., Johnson, W.L., & LaBore, C. (2003b). An interactive pedagogical drama for health interventions. *Proceedings of AIEd 2003*, in press. Amsterdam: IOS Press.

Porter, T. (1997). Creating lifelike characters in *Toy Story*. *SIGART Bulletin* 8, 10-14.

Swartout, W., Hill, R., Gratch, J., Johnson, W.L., et al. (2001). Toward the Holodeck: Integrating graphics, sound, character, and story. In J. Müller, E. André, S. Sen, & C. Frasson (Eds.), *Proc. of the Fifth Intl. Conf. on Autonomous Agents*, 409-416. New York: ACM Press.

Rist, T., André, E., & Baldes, S. (2003). A flexible platform for building applications with life-like characters. In W.L. Johnson, E. André, & J. Domingue (Eds.), *IUI 03: 2003 Int'l Conf. on Intelligent User Interfaces*, 158-165. New York: ACM Press.

Stanislavski, C. (1936). *An actor prepares*. New York: Theater Arts Books.

Telford, K.A. (1961). *Aristotle's Poetics: Translation and analysis*. South Bend, IN: Gateway Editions.

Thomas, F., and Johnston, O. (1981). *The illusion of life: Disney animation*. New York: Walt Disney Productions.

Traum, D. & Rickel, J. (2002). Embodied agents for multi-party dialogue in immersive virtual worlds. In C. Castelfranchi & W.L. Johnson (Eds.)., *Proc. of AAMAS*, 766-773. New York: ACM Press.